# THE

# QUARTERLY JOURNAL

## OF

# ECONOMICS

## VOLUME LXIII

# CONTENTS FOR VOLUME LXIII

## THE TAFT-HARTLEY ACT

### SUMMARY

I. Basic American labor conditions and distinctive characteristics of American trade unions, 1. — II. Efforts of the government to protect the right to organize and some of their unintended results, 6. — III. Provisions of the Taft-Hartley Act, 8. — IV. Areas where the law has not brought about important changes, 12. — V. Areas where the law has produced important changes, 21. — VI. Good features of the law, 20. — VII. Provisions of the law which deal with the "right" problems, but which are poorly worked out, 23. — VIII. Bad provisions in the law, 28. — IX. Basic characteristics of a good law, 29.

The Taft-Hartley Act will soon be replaced by another law. It is therefore especially worthwhile to examine the brief experience under it. What is to be learned from this experience? What are the merits and weaknesses of the Act? Which provisions should be kept, which modified, which discarded? What problems dealt with by this Act must be dealt with also by any successor law?

The Taft-Hartley Act, like the Wagner Act before it, attempted to modify habits of behavior which have been deeply rooted in the American community. In order to understand the law and the problems of drafting a successor act, one needs to understand certain characteristic American labor conditions. Five of these conditions are particularly important:

1. The reluctance of American workers to become union members. This has been one aspect of American individualism. It has been reinforced by the rapid growth of population and the rapid expansion of industrial employment (as much as one-third in a decade during a good part of the nineteenth century) which gave individual workers a good chance to advance. Even as late as 1930, less than twenty per cent of the industrial workers were union members. The growth of trade unions has been largely concentrated in a few periods — such

as 1886, 1897–1904, 1916–1920, and 1933 to date — when conditions were unusually favorable.

2. The fierce opposition of American employers to trade unions. The same individualism which has made the American worker hard to organize has led employers to oppose unions with great vigor. Discharge of men for union activity has often been routine business management, and the use of spies and undercover men in unions has been far from unusual. Only those unions survived which defeated the attempts of employers to destroy them.

3. The frequent failure of American workers to maintain a united front during strikes. In America it has been common for some men to refuse to strike, for some to dribble back to work before a settlement had been made, and for some to be strike breakers. This lack of solidarity, another facet of the community's individualism, is also pronounced among employers. When a strong union strikes against a group of enterprises, individual employers have usually been willing to make settlements without consulting the group.

4. The reluctance of workers to pay union dues after their immediate demands have been met. The "free rider" is a well-known problem of the American union.

5. The reluctance of workers to participate in the day-to-day affairs of the union and to take an interest in its ordinary business. The members expect the officers to run the union and do not wish to be bothered by union affairs. They are reluctant to attend meetings. A well-established union in a recent election of national officers had only forty per cent of the members voting, although there were contests for the presidency and for other important offices. The alleged lack of democracy in American unions has long been a favorite criticism of Socialists and left-wingers. As a matter of fact, the members usually get the kind of unions they desire.

These underlying conditions left their mark on American trade unions — on the structure of the trade union movement, on the way in which unions conduct their business, and on the methods and policies which they pursue. Of these characteristics of American trade unions, the following are of particular interest to us:

1. The dominant position of the national unions in the movement. In the beginning, of course, the trade union movement was a "grass roots" affair. The indifference of many workers to organization and the strong opposition of employers, however, gradually caused organizing at the grass roots level to diminish in importance, and led independent locals to join national unions. Nearly all of the union members in the United States belong to large or fairly large national

unions. At present there are about 200 such unions. Britain, by contrast, has many small, independent local or regional unions. With only about half as many union members as the United States, there are in Britain nearly 1000 independent unions, of which about 750 have less than 2500 members each.

2. The dominance of professional officers, especially national officers. This is in large part a result of the desire of the rank and file not to be bothered about union affairs. The presidents of American trade unions do not have as much authority as presidents of corporations, but their authority is great and has been increasing. An examination of the constitutions of 88 A.F. of L. unions and of 36 C.I.O. unions gives the following results: In 36 A.F. of L. national unions and one C.I.O. union, the appointment of national representatives or organizers of the union is the exclusive responsibility of the president; in eight A.F. of L. unions the president may remove any national officer, and there is no appeal from his decision; in thirteen A.F. of L. unions he may suspend and revoke charters without appeal; in twelve A.F. of L. unions and three C.I.O. unions, he may discipline and replace local officers without appeal. In other cases, the president may appoint national representatives, suspend or revoke local charters, and discipline or supplant local officers, with the approval of the executive board. In still other unions these decisions are left to the national executive board, sometimes with right of appeal to the union convention but often without appeal.[1] In a few unions, contests for national office are usual, but these unions are the exception. In a study of 764 elections to national office in seven unions between 1910 and 1941, Professor Taft found that in 634 elections there was only one candidate.[2] Unions quite typically assert broad authority to discipline their members. Among 81 unions, Professor Taft found discipline authorized for the following offences: "slandering an officer," 29 unions; "creating dissension," 15 unions; "undermining the union or working against its interest," 20 unions; "action that is dishonorable or which might injure the labor movement," 25 unions; "circulating written material dealing with union business without permission of the general executive board," 21 unions.[3] The

1. See Slichter, S. H., *Challenge of Industrial Relations*, pp. 108–109, and 181–189; and Philip Taft, "The Constitutional Power of the Chief Officer in American Trade Unions," this *Journal*, May 1948, pp. 459–471.

2. Philip Taft, "Opposition to Union Officers in Elections," this *Journal*, February 1944, pp. 246–264.

3. Philip Taft, "Judicial Procedure in Labor Unions," this *Journal*, May 1945, pp. 370–385. These vague offences were created to help unions protect themselves from employer agents. This is one of the many ways in which the

reluctance of the rank and file to participate in union affairs and the dominance of the officers have been convenient in some cases for Communists or racketeers who wished to control unions.  On the other hand, the dominance of professional officers has not meant that these men typically lacked support of the rank and file.

3.  The absence of class feeling and the lack of strong concern on the part of national unions for the interests of the wage earning class as a whole.  Not only have employees and employers in America been individualistic, but so also have the national unions.  Each has been jealous of its own independence, and unwilling to accept much guidance or restraint from federations, such as the A.F. of L., which seek to represent trade unions as a whole.

4.  Development by the trade union movement of the principle of exclusive jurisdiction.  This principle means that there may be only one legitimate trade union in a given area — an occupation or an industry.  It also means that the trade union movement asserts the right to determine which union is the legitimate one — that is, which union employees shall join.  The preferences of the men do not count.  Obviously, such a principle could not prevail where the trade union movement was a grass roots affair.  It fits a community, such as the United States, where the professional union officers are dominant.  In Britain, where the influence of the rank and file is strong, the trade unions operate on a principle quite the opposite of the principle of exclusive jurisdiction — for in Britain no union has the exclusive right to organize any occupation or industry.[4]

employers' opposition to unions has made for centralization of authority in the hands of the national union.

4. The British Trades Union Congress does not, however, concede to the rank and file the full right to determine what union they will join.  In 1939, the Bridlington Trades Union Congress, in order to control competition of unions for members, revised the "Main Principles" of "good trade union practice" adopted by the Hull Congress in 1924.  The revision included the recommendations that all unions shall:

  (i) Consider the possibility of joint working agreements with unions with whom they are in frequent contact. . .
  (ii) Not accept a member of another union without inquiry.
  (iii) Adopt the Inquiry Form provided by the General Council in the case of all inquiries under (ii) above, . . . .
  (iv) Not accept a member of another union where inquiry shows that the member is (a) under discipline; (b) engaged in a trade dispute; (c) in arrears with contributions.
  (v) Not commence organizing activities at any establishment or undertaking in respect of any grade or grades of workers in which another union has the majority of workers employed and negotiates wages and conditions, unless by arrangement with that union.

The principle of exclusive jurisdiction has not prevented serious conflicts between unions about jurisdiction over men, for, as I have pointed out, American individualism has produced self-assertive unions which do not readily concede the claims of rival unions or accept the decisions of federations, such as the A.F. of L. Nevertheless, until the rise of the C.I.O., the principle of exclusive jurisdiction was fairly effective in preventing the rise of dual unionism in the United States. This is indicated by the fact that only 13 out of 52 unions founded by secession between 1852 and 1933 survived as independent unions. Twenty-three rejoined the parent union, 10 joined other unions, and 6 expired.

5. The use of short-cut methods of organizing. American unions have not always considered it necessary to get the consent of employees before signing an agreement on their behalf with their employer. Strong unions may approach the employer directly and ask him to sign a union-shop agreement for unorganized employees under threat of a boycott or strike. Then the union approaches the men and signs them up. In a community in which employees were easier to organize and took greater interest in union affairs, such organizing methods would not be tolerated by the workers themselves — nor would they be necessary.

6. The use of strikes and boycotts as methods by one union to fight another over members and over jobs, and the willingness of one union to "raid" another. Raids by one union on another do not mean that the principle of exclusive jurisdiction is not accepted — they mean simply that unions have their own ideas about proper boundary lines and are quite ready to enforce their claims by economic means. In a trade union movement in which *either* the rank and file *or* the federations had more influence, interunion warfare would rarely be carried to the point of strikes or boycotts.

7. Reliance upon the closed shop or the union shop. Both the closed shop and the union shop are distinctive American institutions. Occasional use of them is found in other countries, but they are not prevalent except in the United States and Canada. The closed shop and the union shop are devices partly to deal with the problem of the "free rider," partly to prevent employers from disrupting unions, partly to protect unions from raids, and partly to make internal union discipline more effective.

The attitude toward joining a union is somewhat different in Britain from that in the United States. Here unions put strong pressure upon people to join. There one makes formal application for membership.

## II

American labor conditions, and particularly the difficulties of organizing workers, led eventually to the demand by unions that the government protect the right to organize.  The first laws prohibiting employers from discharging men because of membership in unions were held to be unconstitutional.  The War Labor Board during the First World War gave temporary protection to the right to organize, but the Board's principles and decisions were not legally binding. The Railway Labor Act of 1926 protected the right to organize, but only within the railway industry.  With the National Industrial Recovery Act of 1933 and especially the Wagner Act of 1935, however, the government embarked upon the general policy of encouraging organization.  The effect of these acts was far reaching and, in important respects, unintended.

The basic principle of Section 7a of the National Industrial Recovery Act, and of the Wagner Act, was that workers are entitled to be represented by bargaining agents of their own choosing.  This was a radical departure from well-established American trade union principle and practice — because unions had never conceded that unorganized workers were entitled to join any union which they saw fit.[5]  Section 7a of the National Industrial Recovery Act and the Wagner Act were not intended to undermine the principle of exclusive jurisdiction, but that is precisely what they did.  Each Act gave to the employees in the bargaining unit, rather than to the trade union movement, the right to determine what union, if any, should represent the workers.

Government protection of the right to organize caused a temporary revival of grass roots organizing activity.  In the years 1933 and 1934, when the National Industrial Recovery Act was in operation, 18 new national unions were formed.  Eleven of these were dual unions in the sense that they covered areas of employment already claimed by existing unions.  In the two years 1936 and 1937, shortly after the passage of the Wagner Act, 26 new national unions were formed — nine of them by secession from existing unions.  No one of these nine has rejoined the parent union.  When men became relatively easy to organize, the jurisdictional claims of established unions were widely challenged; the American Federation of Labor, com-

5. The National Industrial Recovery Act and the Wagner Act even gave broader freedom of choice to the rank and file than does the practice of the British Trade Union Congress.  Although the British practice concedes the right of unorganized workers to join the union of their choice, it does not concede full freedom of one union to organize groups of workers within which a second union has members.

mitted to defending these claims, was split; and the challenging unions, together with others which had grown up independently, formed the Congress of Industrial Organizations.

The encouragement of organization by the government increased competition between unions for members, and this competition encouraged high-pressure organizing methods. Some employers were confronted with strikes and boycotts designed to compel them to violate the law by ceasing to deal with the certified union and dealing instead with another union which claimed jurisdiction. Inter-union rivalry increased the interest of unions in the closed shop and the union shop as devices to prevent raids by rival organizations. Interest in the closed or union shop was also stimulated by the rapid influx of new members who were not used to paying dues. During the Second World War the War Labor Board, unwilling to grant the demands for the closed or the union shop, developed the maintenance-of-membership clause. This clause required workers who joined the union to keep in good standing during the life of the contract.

The Second World War delayed the emergence of problems created by the rise of unions, but when the fighting stopped, the country suddenly found itself confronted with new and formidable labor problems. Trade union membership had increased about five-fold between 1933 and 1945 and about two-fold between 1940 and 1945. About one-third of the members were concentrated in six large unions — the Teamsters, the Automobile Workers, the Steel Workers, the Carpenters, the Machinists, and the Miners. About seven million jobs were covered by closed-shop or union-shop contracts (about three times the entire A.F. of L. membership in 1933), and four million more were covered by maintenance-of-membership clauses. This meant that large areas of American industry were open only to union members or to persons whom the unions were willing to admit, and that in these areas loss of good standing in the union meant loss of one's job. The most powerful unions are able to cut off the community from essential materials, such as coal or steel, or to shut down the ports of the country, or to deprive cities of essential transportation. In 1946 the time lost by strikes reached an all-time high — 113 million man-days, or three times higher than in the previous year or in 1937, the two worst years up to that time. On four occasions during the year, nation-wide strikes deprived the country of essential commodities and services, and threatened it with disaster. In February and March, 1946, the steel industry was shut down for four weeks; in April, the country was cut off from virtually all supplies of bituminous coal; in May, the railroads were shut down for

nearly two days; and in November, 1946, the bituminous coal industry was again shut down.

## III

Against this background the Taft-Hartley Act was passed. It is a long law, covering twenty-nine pages of eight-point type; and an ambitious one. It must be regarded as an attempt to provide a comprehensive code to cover the multitude of issues and problems produced, on the one hand, by the traditional hostility of American employers to unions and, on the other hand, by the rapid growth and great power of unions and by the recent break-down of some well-established institutions of the trade union movement. Most of the provisions of the law may conveniently be summarized as dealing with five basic subjects:

1. Relations between unions and individual workers.
2.     "        "    unions and unions.
3.     "        "    unions and employers.
4.     "        "    unions and the community as a whole.
5. Administrative arrangements for enforcing the labor policy of the community.

Some provisions of the Act fall under more than one of these headings, and different persons would make different classifications of the several parts of the law. The following is an incomplete summary of its most important provisions.

1. *Relations between unions and individual workers.* These provisions fall under six principal heads:

(a) The Act specifically recognizes the right not to organize and the right not to strike, as well as the right to organize and to strike; and undertakes to protect all of these rights from interference by either employers, unions or the agents of unions. This means that coercive activities by either unions or employers in connection with organizing activities, strikes, or lockouts are made unfair labor practices.

(b) The Act, like the Wagner Act, recognizes the right of employees to choose their bargaining agent; and attempts to protect this right from interference by employers or unions.

(c) The Act attempts to protect the freedom of workers to enter any occupation or industry by forbidding the closed shop, and by permitting the union shop only when approved by a majority of the workers in the intended bargaining unit.

(d) The Act contains a number of provisions designed to make

elections more representative. Some of the provisions merely convert into statutory form practices of the National Labor Relations Board in its later years. Others require changes in practice. The Act directs that professional employees shall not be put in the same bargaining unit as non-professional employees unless such inclusion has been voted by a majority of such professional employees. It denies the right to vote to "economic" strikers who are not entitled to reinstatement. It provides that the ballot in run-off elections shall be limited to the top choice and the runner-up on the first vote — thereby assuring (contrary to Board practice) that the choice of "no union" appear on the run-off ballot in the event that "no union" was the first or second choice in the inconclusive election.

(e) The Act attempts to protect the right to join a union by forbidding employers from discriminating, in hiring, against employees who are denied membership in a union on the same terms generally available to other members.[6] It also forbids unions which have union-shop contracts to cause employers to discriminate against employees who have been denied membership (except for failure to tender the regular initiation fees and dues), or to collect initiation fees which the National Labor Relations Board "finds excessive or discriminatory under all the circumstances".

(f) The Act attempts to protect the right to remain in a union, by forbidding employers who have negotiated union-shop contracts to discharge employees because of loss of good standing in a union (except for failure to pay regular union dues), and by forbidding unions to cause employers to discriminate against employees who have lost their membership for some reason other than failure to pay the regular dues.

(g) The Act forbids the National Labor Relations Board to consider representation or unfair labor practice cases which originate with unions that do not provide their members with certain financial reports.

2. *Relations between unions and unions.* The Act attempts to deal with conflicts between unions over either jobs or members. It makes it an unfair labor practice for a union to use the strike or boycott (1) to require an employer to assign particular work to employees in a particular labor organization or in a particular trade,

6. Section 8 (a) (3), while permitting employers to enter into union shop contracts under certain conditions, states that "no employer shall justify any discrimination against an employee for non-membership in a labor organization (A) if he has reasonable grounds for believing that such membership was not available to the employee on the same terms and conditions generally applicable to other members..."

craft, or class, unless the employer has failed to conform to an order of the Board determining. the bargaining representative for the employees doing such work, or (2) to compel an employer to recognize the union as a bargaining organization when another union has been certified. Refusal to manufacture, transport, or handle goods, for the purpose of forcing an employer or another person not to handle the products of any other producer, is made an unfair labor practice, and also a basis for the recovery of damages by injured parties. This provision forbids strikes or boycotts against one employer for the purpose of controlling the union affiliation of the employees of another.

3. *Relations between unions and employers.* Some of the provisions of the Act which pertain to relations between unions and individuals or between unions and unions also constitute important regulations of relations between unions and employers. In addition, some of the other regulations of relations between unions and employers are:

(a) The exclusion of supervisors from the protection of the Act.

(b) The broadening of the right of employers to express themselves about unions and of unions to express themselves about employers, provided the statements contain no threat of reprisal or promise of benefits. Many people have assumed that this change was simply in the interest of employers, but it may limit the effect of some sections of the Act upon the activities of pickets.

(c) The authorizing of employers to ask for an election to determine whether a union which seeks bargaining rights represents the employees.

(d) The imposing upon unions as well as employers of the obligation to bargain in good faith, provided the union is the certified representative of the employees.

(e) Provisions making it an unfair labor practice for a union to cause or seek to cause an employer to pay for services which are not performed, or are not to be performed.

(f) Restrictions on financial contributions by employers to representatives of their employees. There are exceptions to the general prohibition. For example, check-off of union dues is permitted on written assignment of the employee and so also are contributions to health and welfare funds, provided the contributions are held in trust for the payment of specified types of benefits to employees, and provided employers and employees are equally represented in the administration of the trust fund, with an impartial umpire to decide disputes.

(g) Provisions authorizing suits in the federal courts for the violation of contracts between an employer and a labor organization without respect to the amount in controversy or the citizenship of the parties.

(h) Provisions modifying the usual rule of agency, by providing that any determination of responsibility of one person for the acts of another, shall not be controlled by the question of actual authorization or subsequent ratification. The Norris-LaGuardia Act went to the other extreme, by making unions and their officers immune from the illegal acts of their members and officials "except upon clear proof of actual participation in, or actual authorization of, such acts, or of ratification of such acts after actual knowledge thereof."

4. *Relations between unions and the public.* In three principal respects the law deals with certain public interests in the activities of unions. It endeavors to discourage Communist leadership in unions by prohibiting the National Labor Relations Board from handling representation cases or unfair labor practice cases originating with a union unless each officer has sworn that he is not a member of the Communist Party or affiliated with it, and is not a member or supporter of any organization that believes in or teaches the overthrow of the government by illegal measures. The Act attempts to deal with the problem of political expenditures by unions, through extending the Federal Corrupt Practices Act of 1925 to apply to trade unions as well as to corporations. Finally, the Act provides arrangements by which the government may deal with strikes or lockouts which threaten to imperil the public health or public safety. Strangely enough, until the Act was passed, the President had no authority to deal with such emergencies, except temporary authority given by war-time legislation, and special authority given by the Railway Labor Act.

5. *Administrative arrangements.* The most important administrative change is the separation of the investigating and prosecuting functions of the National Labor Relations Board from its judiciary functions. This is accomplished by making the general counsel an appointee of the President instead of the Board, placing the regional offices of the Board under his direction, and making him responsible for investigations and prosecutions. Other administrative changes include the requirement that the proceedings of the Board and its agents shall be guided, in so far as practicable, by the rules of evidence followed by the federal courts in civil cases, and the provision of slightly broader powers for the courts to review appeals from the Board. Of considerable importance is the authority given the

general counsel to seek injunctions in some cases of apparent unfair labor practices, and the requirement that he seek injunctions in other types of cases — those in which a secondary boycott has been organized.

## IV

What have been the results of the Act? The law has operated only in a period of high employment — hence the experience under it has limited significance. Certain principal facts, however, stand out. The results may be conveniently reviewed under two main headings: (1) areas where the law has not brought about important changes or where the changes have been temporary only, and (2) areas where the law seems to have produced important changes. Let us first consider where the Act has had little effect, or only temporary effects.

1. The Act has had little effect upon trade union membership. Between 1937 and 1945, membership more than doubled. Hence, a slowing up was expected. This began in 1945. The slow rise which has been going on since 1945 continued after the passage of the Taft-Hartley Act. In the fall of 1948, trade union membership in the United States was approximately 15,070,000, in comparison with 14,841,000 in the fall of 1947.[7]

The Taft-Hartley Act interfered temporarily with the organizing

7. These estimates are derived as follows:

|  | 1947 | 1948 |
|---|---|---|
| A.F. of L. | 7,578,000 | 7,220,000 |
| C.I.O. | 6,000,000 | 6,000,000 |
| Independent unions | 1,836,000 | 2,450,000 |
|  | 15,414,000 | 15,670,000 |
| Less Canadian members | 573,000 | 600,000 |
|  | 14,841,000 | 15,070,000 |

The figures for 1947 are the estimates of the U. S. Bureau of Labor Statistics in Release LS-48-1865. The figures for the A.F. of L. in 1947 and 1948 are based upon the number of members on whom the affiliated unions paid a per capita tax. The actual membership may have been slightly larger. The drop in the membership of the A.F. of L. between 1947 and 1948 is entirely the result of the withdrawal of the United Mine Workers on December 12, 1947. In 1947, the United Mine Workers paid a per capita tax on 600,000 members. The withdrawal of the United Mine Workers also accounts for most of the increase in the membership of independent unions. In both 1947 and 1948, the C.I.O. has claimed 6,000,000 members, but it makes no detailed report of membership.

The deduction for members in Canada is an estimate of membership in Canada of international unions whose headquarters and principal activities are in the United States. According to the *Thirty-Sixth Annual Report on the Labor Organizations in Canada*, published by the Canadian Department of Labor, 573,258 Canadian workers were members of "international" unions in 1946.

activities of unions because it compelled the national representatives to devote much time for several months to union-shop elections. Not all of this time and effort, however, was lost from the union point of view, because the union-shop campaigns helped unions to sign up new members and to bring delinquent members up to date on their dues. Some union leaders blame the Act for the disappointing results of the southern organizing drive. The unions might wisely have postponed the drive until after the union-shop elections had been held, and after the public relations of unions had had an opportunity to recover from the unfavorable effects of the great strikes of 1946. The Act can hardly be blamed for the disappointing results of a drive which was poorly timed and which was conducted with inadequate personnel.

Is the fact that the Taft-Hartley Act had no observable effect upon trade union membership attributable to the strong sellers' market for labor which has existed since its passage? I think not. There is no reason to expect it appreciably to limit the ability of unions to gain members, even in a buyers' market. It contains the same protections against discrimination and economic coercion by employers which are found in the Wagner Act. It does outlaw short-cut organizing methods (such as attempts to force men to join by refusing to handle their products or to service the plants where they work), but these methods have never accounted for a large part of the gains in union membership. The Act gives employers more free-dom to talk against unions, but this is a result difficult to criticize. Union leaders have usually expressed themselves quite frankly about employers.

2. The Act temporarily interfered with the attempts of unions to win recognition through representation elections, but this effect appears to have been the result of special conditions. Some unions could not decide immediately whether to qualify under the Act. Furthermore, there was delay in ascertaining what unions needed to do in order to qualify. Finally, the personnel of unions and of the National Labor Relations Board was busy for months with union-shop elections. For these reasons representation elections dropped by more than half in the year ending June 30, 1948, as compared with the preceding year, and petitions for such elections dropped by almost as much.[8] This was the first drop since 1940 in

8. In the year ending June 30, 1947, 10,677 petitions for representation elections were filed, and 6,920 elections were held; in the year ending June 30, 1948, 6,579 petitions for elections were filed, and 3,211 were held.

Fears have been expressed that employers could get elections postponed indefinitely by trumping up unfair labor practice charges against unions , thus

the number of petitions filed, and the first drop since 1943 in the number of elections held. As the year advanced, the number of petitions and elections rapidly increased. In the last quarter of the year ending June 30, 1948, there were 2,483 petitions for representation elections (virtually the rate of 1947), and 1,261 elections were held.

3. The Act has not affected the percentage of elections lost by unions. This percentage (27.3 for the period August 22, 1947 to June 30, 1948) was the highest since 1936, but the Act has had nothing to do with this figure. There are cycles in the public relations of unions. For example, the proportion of elections lost was at an all-time low in 1937, when it was only 5.7 per cent. It rose gradually to 23.1 per cent in 1939, dropped slowly to 13.7 per cent in 1942, and has been rising ever since. There are similar cycles in the percentage of votes against unions — except that the anti-union peak (24.2 per cent) occurred in the year ending June 30, 1946, undoubtedly reflecting the great strikes of that year. In the year ending June 30, 1948, the percentage of votes against unions had dropped to 23.1.

4. The Act caused a temporary drop in the number of unfair labor practice cases filed against employers. Between August 22, 1947, and June 30, 1948, the number was 2,548 in comparison with 4,232 in the previous fiscal year and 3,815 in 1945–46. The drop was due partly to the fact that some unions had not complied with the law, and partly to the preoccupation of unions with union-shop elections. In the six months ending October 31, 1948, 1,694 complaints were filed against employers. This was almost at the 1945–46 rate. The drop in complaints against employers originating with unions was partly offset by a sharp rise in complaints originating with individuals. There were 1,174 of the latter during the year, but they declined rapidly as the unions themselves became able to file complaints. During recent years the number of unfair labor practice cases filed has fluctuated. It reached an all-time high in the fiscal year 1937–1938 when 6,807 cases were filed, declined steadily to 2,427 in the fiscal year 1944–1945, and increased in the first two years after the war.

5. As a general rule employers have not refused to deal with non-complying unions. This is notably true of enterprises which have contracts with the United Steelworkers, the United Miners, the United Electrical Workers, and the International Typographical

taking advantage of the practice of the National Labor Relations Board against holding elections while charges are pending. See Archibald Cox, "Some Aspects of the Labor-Management Relations Act, 1947," *Harvard Law Review*, November 1947, p. 34.

Union — the principal non-compliers. The attempt of the Waterfront Employers' Association of the Pacific Coast to force the leaders of the International Longshoremen and Warehousemen's Union to take the anti-Communist oath was defeated after a long strike.

6. Hiring arrangements in previous closed shops throughout the country have been little changed by the law. Unions and employers have adapted themselves to the prohibition of the closed shop in various ways. In some cases a union shop has suited both parties as well as a closed shop. In these cases and in others where the union shop existed, the union sought union-shop elections. The Board held an enormous number of these elections — 17,790 between August 22, 1947, and June 30, 1948. The union shop won in nearly all cases with a total vote of 6,164,322, and an opposition vote of 93,938.

The union-shop elections have covered less than half of the employees working under closed-shop or union-shop contracts. In the construction industry, the great stronghold of the closed shop, there was only one union-shop vote — largely because of the virtual impossibility of determining who is eligible to vote in that industry. In other cases unions and employers have worked out ways of evading the closed-shop prohibition. These include mutual discharge provisions which permit unions to require the discharge of "disharmonious" persons, special seniority arrangements which assure that preference in hiring will be given to men of long service in the industry, probationary clauses which permit the employment of a new worker to be terminated by either an employer or the union without an appeal under the grievance machinery of the agreement, and clauses permitting employees to "refuse to continue to work with any employee of the employer whose continued employment is a source of friction among the employees of the employer". In some instances arrangements were worked out by which employers agreed that non-union men would not be hired unless they passed an examination for "competency" given by a joint employer-union committee.[9] In the main, however, employers and unions have met the anti-closed-shop provision of the law simply by ignoring it.

This part of the law imposes no penalties and is not brought into operation unless some one charges an unfair labor practice. Hence, unions and employers can easily continue previous arrangements as if no law had been passed.[1] In some cases there was doubt as to

9. This arrangement, originally suggested by a representative of employers in the printing industry, has since been held discriminatory by the courts.

1. In a large Middle Western city a union in the printing trades and the employers adopted the following clause: "Both parties have been operating for

whether the employer's business was sufficiently in interstate commerce to come under the Act. In other cases, such as the bituminous coal industry and the Pacific Coast stevedoring industry, the Act has been deliberately defied by continuation of arrangements which conspicuously violate the law — a blunt challenge to the government to enforce the law, if it dares.[2]

7. The Act has had little effect upon discriminatory or burdensome impediments to union membership.

8. Unions have not been deluged with damage suits because of violations of contracts or other violations of law. The number of damage suits filed against unions is not known, but the Joint Committee on Labor-Management Relations reports that, up to near the end of 1948, 32 suits had been brought against unions by employers, 19 against employers by unions, and one by an employee against an employer. This is not surprising because the law in reality makes virtually no change in the right of employers to sue unions for breach of contract. Employers have long had this right in the courts of the United States and in nearly all of the states. The Taft-Hartley Act merely removes the previous restrictions that the suit must be for more than $3000 and that in the federal courts there must be diversity of citizenship. Employers have not used the right to sue unions because managements which wish to get production from a group of men do not find it helpful to engage in law suits with the bargaining representatives of these men.

9. The anti-make-work provision of the Taft-Hartley Act has had little effect and, if kept in its present language, is not likely to have much effect in a new law. It is easily evaded. The problem is important because unions in declining industries or in industries where employment is intermittent have a strong tendency to develop make-work rules which tend, of course, to keep down the standard of living of the country.

In some areas the Taft-Hartley Act has produced important effects. Twelve of the consequences of the Act seem to me to be particularly significant:

1. The Act has greatly invigorated the trade union movement.

nearly 50 years under contractual relationship and the traditions and understandings they have developed and are prevailing shall continue as far as possible within the laws."

2. Twelve states have declared closed-shop agreements to be illegal, but virtually no effort is made to enforce the laws. Some state anti-closed shop laws contain no sanctions.

Both the Wagner Act and government controls during the war tended to diminish the self-reliance of the unions. Furthermore, the preferred position of unions in the administration for fourteen years caused them to forget that they could incur the anger of the community and that it might have serious consequences for them. The Taft-Hartley Act has gone far to make union leaders aware that they must not lean too heavily on the government, and that they must avoid too flagrant a disregard for the welfare and convenience of the community. The Act has not had time to produce much effect upon organizing methods of unions, but if the rule against coercion in organizing men is kept in the new law, one may safely predict a great improvement in organizing methods — improvement from the standpoint of unions as well as of the community.

2. The union shop elections have demonstrated the overwhelming desire of union members for the union shop, or probably for the closed shop. Furthermore they have increased the power of unions by helping them to gain union shops where previously they had had only maintenance-of-membership clauses.

3. The withdrawal of the protection of the law from foremen has seriously impeded the growth of independent foremen's unions. After conversion to peacetime production, the impetus of foremen to organize substantially diminished. Fear that managements would make arbitrary decisions in selecting foremen for demotion or layoff at the termination of war contracts had stimulated foremen to organize in the later years of the war. Even in the absence of the Taft-Hartley Act, the organization of foremen would have slowed up, but the Act accentuated this tendency.

4. The provision of the law authorizing employers to ask for elections has caused a considerable increase in such requests. The Wagner Act did not prohibit the National Labor Relations Board to hold elections at the request of employers, but it did not specifically authorize such elections. In the first year of the Taft-Hartley Act, 472 requests for elections were received from employers or individuals, in comparison with 158 received during the preceding year and 82 in the year 1945–46.

5. The provision for decertification elections resulted in 154 such elections up to October 30, 1948. In 105 of these cases the bargaining representative was decertified. This record seems to indicate that the law has been useful in eliminating bargaining agents which had ceased to be active or which had lost the confidence of the employees.

6. The Act has produced a moderate number of unfair labor

practice charges against unions. No such charges were possible under the Wagner Act. From August 22, 1947 to October 31, 1948, 1,110 unfair labor practice charges were filed against unions in comparison with 4,930 filed against employers. The number is large enough to indicate that protection against unfair labor practices by unions is needed. Over half of these cases (631) were brought by employers; but 421 were brought by individual workers, and 58 by unions.

7. The provisions for damage suits against unions for breach of contract have led some unions to limit their liability by clauses in their contracts specifically exempting them from liability for strikes which no union officer called or which the union had not ratified. The most famous of these is the clause in the bituminous coal agreement which states that the contract "shall cover the employment of persons employed in the bituminous coal mines covered by this agreement during such time as such persons are able and willing to work". This clause weakens the contract seriously from the standpoint of the public and of the operators.

8. The Act may be credited with the establishment of machinery to settle jurisdictional disputes in the building trades. Its purpose has been to prevent these disputes from going to the National Labor Relations Board. On two previous occasions in the last 30 years the building trades unions set up machinery for adjusting jurisdictional disputes, but in each case it broke down after a few years because some unions refused to comply with decisions which went against them. The third attempt to set up the machinery, directly attributable to the Taft-Hartley Act, is the most promising one that has been made.

9. The provisions forbidding the National Labor Relations Board to hear cases which originate in unions whose officers have not taken the anti-Communist oath has weakened the influence of Communists in the trade union movement. The deterioration of relations between the United States and Russia has had the same effect. Consequently, one cannot determine the precise influence of the Taft-Hartley Act. Some locals in New York and St. Louis have seceded from the United Retail, Wholesale and Department Store Employees' Union, and there have been secessions from other non-complying unions, such as the United Electrical Workers, the United Public Workers, and the Farm Equipment Workers. Aggressive steps against Communist-led affiliates of the C.I.O. were authorized by the convention of the C.I.O. in November 1948.

Opinion concerning the anti-Communist provisions of the law has undergone considerable change. At the outset union officers

were strongly critical of these provisions, but at the convention of the A.F. of L. in Cincinnati, there was sentiment for strengthening the anti-Communist provisions of the law and for applying them to representatives of management as well as to representatives of unions.

10. The Act has greatly stimulated the interest of unions in political action, has caused them to organize more effectively for this purpose, and has led them to make substantial increases in their expenditures for political action. This has occurred in spite of the provision of the Act against these expenditures — indeed, it has occurred partly *because* of this prohibition. The Act is easily evaded by the device of setting up organizations for political "education."

11. The attempt to narrow the area of industrial conflict by making the secondary boycott unlawful has not been thoroughly tested. Between August 22, 1947 and October 31, 1948 a total of 343 charges alleging secondary boycott were filed. Of these charges, 256 were withdrawn, adjusted, or dismissed in the regional offices of the Board. Secondary boycotts have thus far produced few damage suits.

12. Considerable use has been made of the national emergency provision of the law. In six cases boards of inquiry were appointed; in two, a strike vote on the employer's last offer was held (both votes were in favor of a strike); and in two cases, strikes occurred after the machinery of the law had been used. In two other cases, strikes had started or were about to start when the law was invoked.[3] In a part of one of the maritime disputes (the longshoremen and the stevedores on the Pacific Coast) the National Labor Relations Board attempted to hold a vote of the employees on the employers' last offer. The union ordered its members not to vote, and from Canada to the Mexican border not a single ballot was cast.

3. The boards of inquiry were appointed as follows:
On March 5, in connection with a dispute between the Carbide and Carbon Chemical Corporation and the Atomic Trades and Labor Council at Oak Ridge.
On March 15, in connection with a threatened strike of the United Packinghouse Workers against Swift and Company, Armour and Company, Wilson and Company, Cudahy and Company, and Morrell and Company.
On March 23, in connection with an alleged strike in the bituminous coal industry.
On May 18, in connection with a dispute between the Long Lines Department of the American Telephone and Telegraph Company and the American Union of Telephone Workers (CIO).
On June 11, in connection with a dispute between seven unions in the Maritime Industry and a large number of shipping companies and stevedoring employers.
On August 18, in a dispute between the Atlantic Coast Stevedoring Employers and the International Longshoremen's Association (AFL).

## VI

How good is the Taft-Hartley Act? What are its good points and what are its bad ones?

Perhaps the most important generalization about it is that it attempts to deal by statutory law with many issues which the courts have been handling by common-law, and that it provides federal policies in areas where policy has heretofore been left pretty much to the states. It would not be accurate to say that the policy had been laissez-faire — although *in reality* the only effective public policy in much of the area covered by the Act *was* laissez-faire. However, the Wagner Act had replaced laissez-faire in a considerable area also covered by the Taft-Hartley Act. Furthermore, the state courts had been struggling for years with important issues covered by the Taft-Hartley Act — with the problem of what kinds of strikes and boycotts and strike activities were legal and what kinds illegal. The courts, it is true, are not well prepared to deal with these issues and have lacked effective instruments for enforcing the law.[4]

The problems covered by the Taft-Hartley Act are better dealt with by statutory law than by common-law methods; a national policy is preferable to forty-eight separate state policies, and the machinery for enforcement which is available under the Taft-Hartley Act is superior to that available to the courts. In all these respects the Act is an important step forward. One should observe that, although the Act sets up a uniform national policy, it does not in all cases preëmpt the field and require that the federal policy supersede state policy, even where interstate commerce is materially affected. Nevertheless, the foundation for this ultimate step has been laid.

The provisions of the Taft-Hartley Act may be conveniently divided into: (1) those which seem predominantly good; (2) those which deal with the "right" problems, but which are poorly worked out; and (3) those which are unqualifiedly bad. Among the features of the Act which seem predominantly good, nine are of particular importance.

1. Protection of employers against strikes or boycotts designed to compel them to violate the law. Under the Wagner Act an uncertified union might strike or picket an employer to compel him to deal with it rather than with the duly certified union. The employer had

---

4. This has been particularly true of the federal courts after the Norris-LaGuardia Act limited the use of injunctions in labor disputes. As a result of this Act, the federal courts might be unable to issue injunctions against admitted violations of law simply because the violation was occurring in connection with a labor dispute.

no effective redress — although in some states he undoubtedly was entitled to damages against the union.

2. Protection of employees against coercion from any source in the exercise of their rights to organize or not to organize, to strike or not to strike. Until the passage of the Wagner Act the prevention of coercion in connection with organizing campaigns and strikes was mainly the responsibility of local governments. These did not give protection against economic coercion; and their enforcement of the laws against physical coercion was lax. Under the Wagner Act the federal government undertook to prevent economic coercion originating with employers, but not with unions. The Taft-Hartley Act extends the protection to economic coercion originating with unions as well as with employers, and goes even farther to make physical coercion an unfair labor practice. It is a close question whether or not the federal government should undertake some responsibility for the policing of organizing campaigns, strikes, and lockouts. In favor of it is the poor record of the localities in controlling mass picketing and violence. If the federal government limits its protection to economic coercion, that protection should be given regardless of whether the source of economic coercion is an employer or a union. A serious defect of the Act is its failure to define or limit clearly the meaning of "coerce" or "restrain," as used in Section 8 (a) and (b). As a result, too much discretion in defining the words is left to the National Labor Relations Board and to the courts.

3. Comprehensive protection of employees in their right to choose their bargaining agents without interference. The Wagner Act gave only partial protection, for it permitted unions to influence elections by threatening to refuse to handle the product of the plant if the workers chose the "wrong" union.

4. The provision in the law safeguarding freedom of speech. Some think it goes too far, in forbidding the Board to consider as evidence of an unfair labor practice statements which do not contain threats or promises of benefits. The question is a close one, but it is well to decide the doubt in favor of freedom of speech. As I have indicated above, this provision affects unions as well as employers.

5. Changes in the election rules designed to help employees select bargaining agents who represent their real choices. Most of these changes are desirable. Certainly the professional employees should not be put, against their will, in the same bargaining unit as non-professional workers. It is difficult to defend the Board's refusal to include the "no union" choice on the run-off ballot when this choice received the largest or second largest number of votes in an

inconclusive election. Quite controversial is the question as to whether "economic strikers" who have been replaced should be allowed to vote. The Act says not. There is an obvious difficulty in giving votes to men who will never again be employed by the company. Against this view it is argued that denial of voting rights to displaced men will encourage employers to break unions by replacing strikers. The rule of the Taft-Hartley Act seems, on the whole, to be sound, but the danger of abuse exists. The Board can probably prevent abuses by refusing to hold elections soon after strikers have been replaced by non-strikers.

6. Authority for employers to request an election in cases of disputes over representation. This provision may not mean much change in practice because the National Labor Relations Board has in late years considered requests for elections from employers. In the early years, however, it did not. The Act does not require that the Board grant the employers' (or the unions') requests.

7. Removal of the ambiguity found in the Wagner Act as to whether the government's policy of protecting the right to organize applies to supervisors. First the National Labor Relations Board held that the Wagner Act applied to supervisors, then that it did not, and finally again that it did. This question should be settled by legislation rather than by administrative decisions. Many persons think that the Taft-Hartley Act should have included foremen among employees. I believe that the provisions of the law on this point are sound. The great stake of the country in efficient management is better promoted if foremen are not given special encouragement to form trade unions. If conditions in a plant are bad enough to make foremen keenly desire a union, they are in a good position to organize.

8. The requirement that unions as well as employers bargain in good faith. The provision will probably have little effect because unions which are strong enough to refuse to bargain on certain subjects are also strong enough to reject successfully all proposals which employers may make with respect to these subjects. Nevertheless, the development of good industrial relations was hindered when employers found themselves compelled to bargain on any proposal made by a union, and yet were unable to get the union even to discuss certain matters which were covered by its constitution.

9. Limitations on the financial payments by employers to trade union representatives. These limitations, which carry criminal penalties, are a useful protection against extortion and "shakedowns". The regulation of so-called health and welfare funds is needed to

prevent abuses of these funds and to keep union politics from affecting their administration, since thousands of employees rely on these funds for a substantial pension which will not become payable until after twenty or thirty years. On the other hand, the Taft-Hartley Act does nothing about the highly dubious method of financing welfare funds by imposing a private "sales tax" on coal, records, or other products.

## VII

Some provisions of the Taft-Hartley Act which attempt to deal with the "right" problems — that is, important problems which need attention — are poorly worked out and will not accomplish the intended results.

1. Provisions designed to protect the right to join a union. If the closed shop or the union shop is permitted (and they fit American conditions), the right to join a union needs to be safeguarded. It will be recalled that in 1946 about seven million jobs were covered by closed shop or union shop contracts. Organizations which seek to determine who may pursue a trade or who may enter an industry are no longer purely private, and their admission requirements are affected with a public interest. Keeping the doors of unions open is no academic matter. About 20 unions have provisions in their constitutions against the admission of negroes, and other unions practice discrimination without requiring it. A few decline to admit women, and many impose high initiation fees. Among 300 locals affiliated with A.F. of L. unions, Professor Taft found that 65 charged initiation fees of $100 or more, and 24 charged fees of $150 or more.[5]

The provisions in the Act designed to protect the right to join a union are unsatisfactory. As I have pointed out, the Act forbids employers to justify discrimination against employees for non-membership in a union on the ground that the union itself discriminates, and makes it an unfair labor practice for a union "to cause or attempt to cause an employer to discriminate against an employee" because of membership or non-membership in a trade union. These provisions presuppose that the initiative in enforcing the Act will come from employers. Few employers are likely to incur the risk of a strike or a lockout by challenging the admission policies of a union, because employers usually have little or no stake in unions' refraining from discrimination.

Likewise of little use is the declaration that it is an unfair labor

5. Taft, Philip, "Dues and Initiation Fees in Labor Unions," this *Journal*, February 1946, p. 222.

practice for unions to require employees covered by a union shop agreement to pay an admission fee in an amount which the Labor Relations Board "finds excessive or discriminatory under all the circumstances". Complaints are not likely to be initiated either by individuals or by the Board. Furthermore, no definite yardstick is given to the Board as to what fees are excessive. The federal government would do well to set a maximum — say $75 or $100. These amounts would protect men in intermittent trades from losing employment to persons who do not regularly pursue the trade, and yet would be small enough not to be a severe hardship to persons who wished to enter the trade permanently.

2. Provisions designed to protect the right to remain in a union. The union shop, the closed shop, or the use of maintenance-of-membership clauses, require that the right to remain in a union be protected; but the provisions in the Act on this point are unsatisfactory. The law makes it an unfair labor practice for the employer to terminate the employment of a worker covered by a union shop contract if the employer "has reasonable grounds for believing that membership was denied or terminated for reasons other than the failure of the employee to tender the periodic dues and the initiation fees uniformly required as a condition of acquiring or retaining membership". The law also makes it an unfair labor practice for a union to cause an employer to discriminate against an employee whose membership in the union has been terminated on some ground other than his failure "to tender the periodic dues and the initiation fees uniformly required as condition of acquiring or retaining membership". This is too narrow. There are many offenses for which a union may wish to discipline a member, and if he can leave the union and still hold his job, the union loses the chance to impose effective discipline. A better arrangement would permit unions which have union-shop or closed-shop contracts or maintenance-of-membership clauses to require employers to dismiss men who lose their good standing, regardless of the cause; but would require also that the worker be permitted to continue to hold his job while exhausting his rights of appeal within the union, and would provide for an appeal from the highest union authority to a neutral agency — preferably the National Labor Relations Board.[6] If a man lost this appeal, the employer should be required to dismiss him. If he won, he should be permitted to keep his job, but the union should be permitted to exclude him from membership.

6. Of considerable importance is the status of a union member during appeal from a penalty imposed by a local union. Professor Taft found five out of 126 unions in which the penalty is vacated during appeal, nine in which the

3. Provisions designed to discourage Communist control of trade unions. The Communist problem in American trade unions, where actual Communist membership is very small, is a reflection of the unwillingness of American wage earners to concern themselves with union affairs. The use which Communists have made of trade unions in other countries indicates that the discouragement of Communist control of unions is an appropriate object of public policy. It is reasonable to require the officers and key employees of unions to declare (under penalty of perjury) whether they are or recently have been members of the Communist Party. Both union members and employers are entitled to know this. The Taft-Hartley Act does not go far enough when it permits union officers and key employees to refuse to declare whether or not they are Communists; but it goes too far when it deprives non-complying unions of the benefits of the law — although the latter disadvantage is diminished by the opportunity of individuals to initiate complaints. It is not sound public administration to make the benefits of a law available only when union officials have taken a given oath because the rank and file frequently have little control over their officials. Some people, however, believe that denial of the benefits of the law to Communist-led unions is necessary in order to force the members to take sufficient interest in union affairs to rid themselves of Communist officers.

4. Provisions designed to control the political expenditures of unions. Some control of the political expenditures of unions is sound in principle. Unions are essentially bargaining organizations, hence they need to admit all those in an occupation or industry regardless of political preferences. There are obvious objections to allowing organizations which impose union-shop or closed-shop contracts to collect money from Republicans to help elect Democrats or from Democrats to help elect Republicans. The provisions of the Taft-Hartley Act designed to limit the political expenditures of unions are ineffective because, as I have pointed out, they can be evaded by calling the activity "education". I know of no way of making these provisions effective. Nevertheless there is some merit in applying the same restrictions on political expenditures to both unions and corporations.

penalty imposed by a local is not enforced in event of appeal until it has been approved by a general officer or the general executive board (which may not be the ultimate appeal body), 52 in which appeal does not suspend the verdict, 42 in which the status of the appellant is not defined, and 18 in which miscellaneous arrangements prevail. Philip Taft, "Status of Members in Unions during Appeal from a Penalty Imposed by the Local Union," this *Journal*, August 1948, pp. 610–616.

5. Provisions for dealing with disputes which imperil the public health or safety. The procedures of the Act are initiated when the President believes that a strike or lockout, or a threatened strike or lockout, is likely to imperil the national health or safety. The first step is for the President to appoint a board of inquiry to report on the facts and on the position of the parties. The report must be made public, and the board is specifically enjoined not to make recommendations. The appointment of such a board does not affect the right of the parties to initiate a strike or lockout. On receiving the board's report, the President may direct the Attorney General to seek an injunction enjoining the strike or lockout. If the district court finds that the strike or lockout would imperil (or does imperil) the national health or safety, it may issue the injunction, and the board of inquiry is thereupon reconvened. If the dispute is unsettled at the end of sixty days, the board reports to the President the current position of the parties and the efforts which were made to settle the dispute. This report must also be made public. Within fifteen days after the second report, the National Labor Relations Board is directed to take a secret ballot of the employees of each employer involved in the dispute on the question of accepting the final offer of the employer. Within five days after the vote the National Labor Relations Board must certify the results to the Attorney General who must then move for discharge of the injunction. The court must grant the motion. The parties are then free to initiate a strike or lockout. The President, however, is directed to make a full report of the proceedings to Congress.

Certainly the provisions for obtaining a postponement of a strike or lockout which would imperil the national health or safety are reasonable. But the other provisions for dealing with national emergencies are seriously defective.

To begin with, the arrangements create the danger that two rival agencies will simultaneously be attempting to settle the same dispute. Thus the Act plainly contemplates that the Conciliation Service will be attempting to settle it while the board of inquiry is ascertaining the facts; but it leaves the role of the board uncertain, since it does not forbid conciliation by it. Nothing is more likely to prevent successful conciliation than rival efforts to settle a dispute. Under the Railway Labor Act, by contrast, the so-called emergency boards, which are similar in some respects to the boards of inquiry authorized under the Taft-Hartley Act, do not come into existence until the mediators have formally withdrawn from the case.

If boards of inquiry are used, they should be made subordinate

to the Conciliation Service — that is, appointed only at the recommendation of the Director of Conciliation — and they should either be prohibited from conciliating or not be permitted to come into existence until after the Conciliation Service has confessed failure and dropped the case. If boards of inquiry are made a second step in handling disputes which create national emergencies, the prohibition against their making recommendations should be removed.

The worst features of the Act for dealing with so-called "emergency" disputes come into operation if the dispute is still unsettled after the injunction has been in effect sixty days. The National Labor Relations Board must then take a vote of the employees on the employer's last offer. The framers of the Act apparently assumed that this vote would be a genuine expression of the willingness of the employees to strike, and that the employees would terminate the dispute by accepting terms which their leaders had rejected. There is no evidence to support this naïve notion. Experience under the Smith-Connally Act during the war indicated quite the opposite: obviously men will not desert their leaders at a time of crisis for the union. In two votes held under the Act, the union members stood behind their leaders. Futhermore, in one of these cases, the members even rejected an offer which had been made subsequent to the "last" offer, and which their leaders had accepted. In a third case the men, under instructions from the union, refused to vote.

The provision that the government conduct a vote on whether or not the men shall imperil the national health and safety puts the government in a ridiculous position. The vote implies that a strike is permissible provided only that the men authorize it; hence the government is tolerating an attack on the community's health or safety. Obviously the only role which the government can decently play is one of defending the community — not of sanctioning strikes against it.

Particularly bad is the provision that after the vote has been taken, the injunction shall be vacated. Consider the psychological effect of the sequence of events — a vote rejecting the employers' last offer immediately followed by vacating the injunction which up to that moment has prevented the strike! Certainly these arrangements are well designed to convince the men that they have met all conditions and are now free to strike regardless of the consequences. As a result, the dispute has been made more difficult to settle. At precisely this point the Act stops. It provides no additional steps by which the dispute may be postponed or settled, stating merely that the case may be referred by the President to Congress, in case

Congress happens to be in session. This is an ideal way of getting industrial relations into politics.

It is obvious that if interruption of production or service would imperil the national health or safety, the government must have authority to assure that production or service continue — at least in amounts required to prevent disaster. One arrangement might be to authorize the President to require the parties to observe the recommendations of the board of inquiry (this assumes that the board is permitted to make recommendations) for a limited time — say six months, or long enough for the parties to make a fresh start on bargaining over their differences; another arrangement might be to authorize government seizure and operation of the enterprise; still another might be to authorize the President to arrange for a limited amount of service at the previously existing terms of employment subject to retroactive adjustment when the dispute is settled.

6. The provision authorizing the General Counsel to apply for injunctions to stop apparently unfair labor practices, pending final determination of the issue by the board. There are likely to be many cases in which justice is best advanced by maintenance of the *status quo* pending a decision on the legality of a given action. Only the General Counsel may seek injunctions, and the court, of course, may not see fit to grant his petition. As of November 1, 1948, six permissive applications and thirty-one mandatory applications had been made. Two of the discretionary applications were against employers. As a result of one of these the General Motors Corporation was restrained from installing an insurance plan without negotiating with the union.

Although the provisions for injunctions improve the enforcement machinery, they have two important defects. One is the requirement that the General Counsel seek an injunction in certain types of cases (secondary boycott). Surely this should be left to his discretion. The other is an ambiguity as to whether the General Counsel must show that the practices which he would restrain would cause irreparable damage and cannot adequately be remedied at law. One district court has held that these traditional standards in granting injunctions do not apply.[7] Yet there seems to be no good reason why they should not.

## VIII

Of the features of the Act which are completely bad, the two most important are the prohibition of the closed shop and the

7. *Dowd's* v. *Local 294, Teamsters Union.* 75 F. Supp.( N.D. N.Y.)

requirement that union-shop contracts be approved by a vote of the employees.

The closed shop is a well-established American institution which fits conditions here and which serves useful purposes — especially in occupations where much of the work is intermittent. It creates problems and is subject to abuse. Nevertheless, public policy should aim at making the closed shop work better, not at abolishing it. The fact that the Railway Labor Act imposes criminal penalties on managements which sign closed-shop (or even union-shop) contracts is not in point. The closed shop has never existed in the railroad industry, and the anti-closed-shop provision of the Railway Labor Act was actually put in at the demand of the unions — although they now demand its removal.[8]

It was a mistake to require that union-shop contracts be approved by a vote of the employees. Such votes serve no useful long-run purpose. Employers are not willing to sign closed-shop or union-shop contracts unless they believe that their employees strongly desire such a contract. Since managements are not easily convinced of this fact, there is little danger that closed-shop or union-shop contracts will be imposed upon unwilling workers. In a few industries, such as the building trades, compliance with the law is virtually impossible because no one can determine who is eligible to vote.

## IX

What kind of a labor law should the country have? It would not be appropriate in this article to attempt to spell out the details of a successor to the Taft-Hartley Act. The comments already made have given some indications on particular matters, and a few broad generalizations may now be set forth.

1. The law should be comprehensive. This does not necessarily mean that it should be long and detailed. Nevertheless, there is a wide range of problems on which public policies are needed — problems of relations between unions and individual workers, between unions and employers, between unions and unions, and between unions and employers on the one hand and the community on the other.

2. It should take account of the fundamental characteristics of industrial relations in the United States — those which are the product of the American environment and which cannot easily be changed. This does not mean that no attempt should be made to

8. Since some of the railroad unions discriminate strongly against negroes, the closed shop should not be permitted on the railroads until negroes are adequately protected against discrimination.

change them, but the principal objective should be to protect the essential institutions of American industrial relations against abuse, and to make them work more satisfactorily. For example, it would be a mistake for public policy to attempt to undermine the dominant position of the national union in the American trade union movement, or to attempt to convert the movement into a grass roots affair in which union officers have much less influence and the rank and file more influence than is now the case. The strong national union fits American conditions, and American employees wish their officers to run the unions. These conditions, to be sure, create certain problems. Public policy should attempt to deal with these problems rather than to change the nature of the trade union movement itself.

3. It should attempt to narrow the area of industrial conflict. Certain uses of strikes and boycotts are not appropriate, and it is high time that trade unions and the community dropped the naïve notion that the unlimited right to strike is an inherent and necessary right of free men. In general, it is inappropriate that the strike should be used by unions to battle other unions, or that strikes or lockouts should be used to coerce the government or to compel changes in public policy.

4. It should attempt to deal with certain specific problems. The following six are particularly important:

*a.* It should protect the right to organize. If this right is not protected, employers will use discharge and other forms of economic coercion to break up new unions. Protection of the right to organize should include protection against economic coercion by unions as well as by employers. Otherwise, some unions will attempt to control men's choices by boycotts, strikes, or other short-cut methods. There was a good case for tolerating such methods so long as employers were free to discharge men for joining unions. But now that coercion by employers is unlawful, public policy should also prohibit the use of economic coercion by unions to force men to join or to leave a union. It is anomalous for the legality of economic coercion to depend upon who uses it.

*b.* It should protect the right of workers to pick their bargaining agents free of interference from any source. The Wagner Act destroyed the effective operation of the principle of exclusive jurisdiction, and a new arrangement must be developed to take its place. The British principle of tolerance will not do because it presupposes more influence of the rank and file in the trade union movement than has been traditional in the United States. The trade unions themselves are not capable of providing a solution because the largest and most powerful unions, which do not wish interference from any source,

have too much influence within the labor movement. Hence, inter-union rivalry over members can be controlled only by public policy. The very growth of unions makes such control imperative. When two powerful unions battle for the control of workers, as did the carpenters and lumbermen in the Northwest several years ago, or as have the teamsters and brewery workers in some cities, or the long-shoremen and the teamsters, the costs to the community may be intolerable.

*c.* It should permit closed-shop contracts where the union represents the majority of the employees in the bargaining unit. The federal government would do well to preëmpt this field and to forbid states from outlawing closed-shop contracts in plants engaged in interstate commerce.

*d.* It should safeguard the right to become a member of a union and the right to remain in a union, wherever closed-shop or union-shop contracts exist.

*e.* It should make strikes over the jurisdiction of unions an unfair labor practice, unless the employer has denied work to members of the union which has the bargaining rights for the job.

*f.* It should safeguard the health and welfare funds which are being set up, and which are likely to be set up in increasing numbers. It would be desirable to reduce the incentive to establish such funds by greatly increasing the tax base and benefit base in the government scheme of old age and survivors' insurance. The pension which a man receives should not depend too much upon whether he happens to work for a generous employer, or to belong to a strong trade union. Furthermore, private health and welfare funds are likely to reduce the willingness of employers to hire men above forty-five, and thus seriously to limit the mobility of labor.

*g.* It should provide arrangements for handling disputes which imperil the national health and safety.

Almost as important as the provisions of the law are the way in which it is drafted and the sponsorship which it receives. A great forward step would be taken in industrial relations if representatives of employers and of unions could reach agreement on the basic provisions of a successor to the Taft-Hartley Act. Unfortunately, the prospects for such agreement are not bright. Some years will probably be required for the country to develop a well-considered and well-balanced industrial relations act. Nevertheless, with the experience of the Wagner and Taft-Hartley Acts as a guide, the new Congress should be able to draft a law which is substantially better than either one of them.

HARVARD UNIVERSITY                                  SUMNER H. SLICHTER.

# BRITAIN'S ECONOMIC PROBLEM[1]

SUMMARY

I. The causes of the British crisis, 32. — II. Pent-up demand and inflation 33. — III. Decontrol, military commitments and the crisis, 37. — IV. The elements of a solution, 50. — V. Conclusions, 60.

## THE CAUSES OF THE BRITISH CRISIS

The overwhelming majority of British economists attributed the British crisis of 1947 to inflation, in its turn due to budget deficit and to the artificial lowering of the rate of interest. In their view this stimulated excessive investment, and also caused a diversion of resources to the production of non-essential goods. It "emptied" the economy of working capital and led to a fall in productivity. The measures taken by the Government — direct controls limiting imports and enforcing higher exports, reduction in investment and increases in taxation — were not considered sufficient to bring about balance. The policy of stabilizing wages and prices was regarded as tinkering with symptoms. The solution of the British problem must, in their opinion, be sought in the restoration of the working of the free-pricing system, the abolition of direct physical controls. This must be coupled with disinflation, i.e., the reduction of global monetary demand by a budget surplus obtained mainly by a severe cut in the subsidies now paid to stabilize the prices of necessities. Equally, investment would have to be reduced by an increase in the rate of interest and, if necessary, by taxing investment. It is recognized that unemployment would have to rise if inflation were to be avoided without controls. It is argued however that the increase in productivity which would result from this policy would much more than offset the loss caused by the decline in employment. The

1. Based on a paper given to the Harvard Economics Club in the spring of 1948: the author is much indebted to the Rockefeller Foundation for enabling him to visit the United States. The opinions expressed are, of course, entirely personal. The fact must be strongly emphasized that the British economic system and social background differ so much from those in the United States that conclusions based upon the analysis of one cannot be applied to the other. Investment capacity in the United States, e.g., is so vast, technical progress so quick, the agricultural manpower "reserve" still so considerable, that the elasticity of the system, even without unemployment, ought to suffice. Unfortunately, little attention is paid to these differences in many of the current controversies.

foreign balance on the other hand should be treated by devaluing sterling.

In the present paper an attempt will be made to show that:

(a) the purely budgetary and investment factors played a much smaller part in the crisis than has been suggested, and could in any case not be controlled by the measures proposed.

(b) the record of the British Government was far more creditable than it is held to be by opinion in this country, and that such relative failure as occurred was due to the premature de-control of important sections of the economy, not so much causing a fall in productivity as preventing a full and quick expansion.

(c) in addition, a lack of effective co-ordination led to commitments which were far beyond the capacity of the country to sustain, especially as the clamor for increased consumption — in which a number of reputable economists participated — did not permit the Government a careful husbanding of the resources set free by demobilization, tardy as it was.

(d) the restoration of the "free" markets would not restore equilibrium internally or externally but would result in social strife and economic relapse.

The striking success of the Government in reducing the international deficit within the limits of the Marshall Aid; the worsening of the international situation; and the consequent danger that productive resources will be diverted towards armaments without a countervailing increase in foreign help, makes this a convenient moment to survey the post-war British economic development. The new phase which opens can no longer be regarded as "post-war." But the experiences of the period immediately following the end of hostilities should be very valuable in framing policy.

## II. PENT-UP DEMAND AND INFLATION

The general tendency to attribute the convertibility crisis in 1947, and the continued grave difficulties of Britain, to the monetary tension, to "suppressed" inflation; and to look for the cause of the latter in an excess of current spendable incomes, and more especially in budget policy, conceals the most important cause of our present problems even in the monetary sphere. It is true that the Government deficit was larger than suggested by the old-fashioned Exchequer statements. These treat certain capital receipts as current revenue, and exclude other current payments, since they are not met through budget votes. The true deficit of the public authorities (central and

local) was some £617 million in 1946 and £135 million in 1947.[2] It must be remembered, however, that the revenue flows most heavily in the first quarter of the calendar year (the British financial year begins on April 1) and that most of these payments are anticipated by the corporate and individual taxpayers in their decisions. If we take this into account the combined account might be said to have shown roughly £400 million of deficit in 1946 and a surplus of some £100–200 in the past year. The gross capital expenditure of the local authorities, however, increased by £122 million in 1946 to £213 million in 1947, which was only partially offset by the excess of sales of assets of the Central Government over investment (£27 and £22 million respectively). The deficit and increase in public investment was certainly below the £500 million mark even in 1946, and in 1947 it was inconsiderable. There was a dramatic fall in the government expenditure on goods and services from £4.5 billion in 1945 to under £2.5 billion in 1946 and £2.2 billion last year. With national income running at the level of some eight billions there is no reason to suppose that it could have been to any considerable extent the cause of excess monetary demand.

The monetary (as contrasted with the production) problem is centered on the weight, and even more on the distribution, of the "pent-up" demand, both for consumers' and for producers' goods accumulated during the war, when both individuals and firms were forced to abstain from spending. It is difficult to measure the volume of that part of the enforced accumulation of liquid resources — by no means restricted to "money" — which was unstable in the sense that it pressed on the market. On the basis of all historical evidence we must expect that a considerable part of these reserves — however "enforced" they had been initially — will "settle down": the demand for liquidity has shown a secular upward trend. But how fast and how far this "settling down" will proceed will depend (1) on the distribution of these reserves and (2) on the confidence in money. A rise in prices and a shift towards profits would — if wage-demands were not pressed successfully in consequence—undoubtedly accelerate the exhaustion of consumers' pent-up demand and the entrepreneurial demand for capital goods can be more easily controlled. But wage demands are likely at full employment if prices rise. And a vicious circle of rising prices and wages might entail a mobilization

2. Death duties and other tax payments out of capital are included in the revenue and are taken into account by diminishing private savings. If we exclude them the deficit becomes £729 and £298 million respectively. Most of the critics overshot the mark by including capital payments in the reckoning of the current deficit.

of quiescent reserves. Thus curing inflation by inflation[3] is not devoid of grave risks, while its "suppression" might in the end remove all need for further "suppression" because an increasing portion of the liquid reserves will gradually be stabilized. Last but by no means least, it should be borne in mind that the "overhang" represented a more volatile potential demand than would an excess of current income due to, say, government deficit. A sudden change in expectations might therefore change abruptly the "monetary" background and result in deflation.[4]

There can be little doubt that in Britain a considerable part of the war-time savings and reserves did press on the market whenever and wherever it was not prevented by direct controls. In the case of consumers the volume of pent-up demand was estimated immediately at the end of the war at some £2,500 to 3,000 million, a multiple of the possible "inflationary" impetus imparted by government finance. A considerable (and the most unstable) part belonging to the lower income classes must have been used and exhausted itself in raising the price of non-essentials and services (including gambling) in the past two years. But at least a third, perhaps more, still "overhangs," mainly concentrated in the hands of the more prosperous working and lower middle-income classes which have greater resistance to high prices. It is this, easily retractable, demand which was mainly responsible for the pressure on the consumers' goods markets. It is still by far the most important potential cause of unsettlement in the markets for consumers' goods, especially durable consumers' goods (including house-repair), the shortage of which is grave and for the purchase of which it has been accumulated.

So far as investment demand is concerned — including working capital consisting of semi-finished or finished consumers' goods — the situation is analogous. The accumulation out of the war-time reduction of stocks, non-invested depreciation allowances, plowed back profits, excess profits tax repayments and war-damage compensation is well over £4,000 million. The retractability of this demand depends even more than the previous category on price-expectations. As long as there is a continuous upward pressure on prices this demand will contribute to it. Once that pressure ceases it might be withheld with grave repercussions on the balance of investments: for any fall in private investment must under present conditions be met by an increase in public investment in housing, etc.; and there can be no

3. As for instance suggested by those who wish to suppress the subsidies stabilizing the prices of necessities.

4. There seems to be mounting evidence that this in fact has been happening since Christmas 1947. Cf. below.

doubt that the latter — for however desirable objects — will not increase productivity as quickly as private investment in fixed capital.

This peculiar nature of the inflationary forces limits the usefulness of most of the measures advocated. Only a complete blocking and a slow release of the financial reserves could restore equilibrium instantaneously. Increased taxation alone would be offset (and might be much more than offset) by dis-saving. But measures such as complete blocking would entail chaos and unsettlement in the financial sphere. The hardship caused would be out of proportion to the advantages gained. A forced loan or a capital levy would also tend to undermine the willingness to hold cash or liquid assets with the attendant grave long-run consequences[5] — unless indeed full-scale State ownership were — as it is not — contemplated. It is very questionable, moreover, whether the smaller savings which are relatively more unstable, and therefore more dangerous, could be immobilized without serious political consequences. And the "freezing" of capital reserves of firms would involve discrimination against the more conservative (and patriotic) firms who have "disinvested" most. Nor can complete financial freezing in any way be regarded as a less comprehensive or administratively easier scheme than the direct control over investment.

Thus it would seem that the possible contribution of budgetary policy to the re-establishment of "equilibrium" is much more modest than suggested by the advocates of "monetary, global controls." This is not to say that a budget surplus should not be aimed at. Over a period of years it would certainly help (together with measures aiming at increasing productivity) in dealing with the huge volume of liquid reserves accumulated as the result of the war. But it would be folly to hope that budget surpluses could in the short run re-establish monetary conditions which would enable wholesale decontrol without inflicting such damage on the financial structure of the country and on the standard of life of the lower income classes as to prove politically intolerable and result in an open inflationary spiral.[6, 7]

5. The "special contribution" introduced in the last budget represents a capital levy based on the income of assets held. It is, from this point of view, not without dangers, especially as voices have already been raised for its repetition as against the assurance of the Chancellor that it will be a unique operation.

6. The example of France and Italy should be a warning. The German currency reform (unless indeed speedily followed by a large-scale capital levy and the equalization of the burden of the war as well as the re-introduction of direct controls) will probably also fail.

7. It will be argued below that a budget surplus should also be aimed at in order to increase savings, i.e., the investment capacity of the country at full

Even less plausible is the argument that dear money would improve our position. Given the huge liquid reserves, the grave scarcity of productive capacity, a small increase would have been completely ineffective in checking investment or concentrating it on *short* term improvements.[8] Most of the "long" capital projects which have been recently under attack have been launched, in fact, mainly on the basis of subsidies (e.g., housing, public improvements) and not on the basis of cheap money. And a large increase in the rate of interest (as in some countries after hyper-inflation to over ten per cent per annum) would have had devastating effects on the budget and solvency. The pursuit of ultra-cheap money was undoubtedly mistaken. But this was not because of its direct effects on the rate or composition of capital investment.

On the other hand, the importance of the implication of full employment and of the strong bargaining position of the trade unions was largely disregarded. As long as national unity is preserved the trade unions could be expected to show, and did show, moderation in pressing for higher wages. But the cruder sort of disinflationary measures proposed, i.e., the abolition of food-subsidies, derationing and decontrol, would have ended "fair shares," and, as on the Continent, unleashed wage demands, thus causing an endless spiral of prices and wages. In the end such a process induces a self-inflammatory explosion. The conscious creation of unemployment to prevent this, however, was not practical politics — in Britain, under a Labor Government; and even under the Conservatives might have ended very differently (even from an economic point of view) from what its partisans seem to envisage.[9]

We must conclude that global monetary measures could not alone have brought about a solution of Britain's troubles or enabled her to dispense with direct controls. It was the lack of co-ordinated planning of the use of physical controls and related matters which resulted in the British crisis.

### III. DECONTROL, MILITARY COMMITMENTS AND THE CRISIS

The tasks before the British Government at the end of the war were truly appalling in magnitude. And the psychological background against which it had to devise its measures could hardly

employment. Otherwise we should have to tolerate an increase in the inequality of income or continue to enforce saving by direct controls increasing the monetary "overhang" of pent-up demand.

8. And, equally, influencing consumers' decisions.

9. Cf. note 6, p. 36.

have been more difficult and unpropitious. (1) The British population undoubtedly suffered less in the last stages of the war than most Europeans. But the war had lasted longer and the need in 1940–41 to resist a much superior enemy alone, combined with the continued shipping crisis, had diminished consumption by some 15 per cent[1] and the period of deprivation had been extended. Thus the population expected an immediate and substantial relief at the end of hostilities.[2] (2) At the same time the state of capital equipment was such as to justify serious apprehension. Britain had entered the war in 1939 with a capital stock which was relatively inferior to that of her main competitors[3] — despite the recovery made after 1932. A number of interesting studies have shown the continued relative decline of British manufacture in comparison with that in the United States and even in Germany. The effects of this decline on the standard of life were largely offset in the inter-war period by a substantial improvement in the terms of trade. During the war, as we have seen, a further serious depletion and obsolescence took place, not least in working capital. In view of the loss of foreign assets and income, however, and the consequent need for increased exports, the improvement in the terms of trade could no longer be counted upon to help the country. The increase in world population and war devastation made a worsening not unlikely. In view, in particular, of the pledge given by the Coalition Government to adhere to "free market" economics on a non-discriminatory basis in international relations, the most determined effort to re-establish the British industrial efficiency was essential, if the country's standard of life was to be safeguarded.

A substantial reserve of productive power was available, of course, in the very fact that Britain had mobilized her manpower (including women) more fully than had any country but Russia. In 1944 the Government expenditure on goods and services amounted to £5.2 billion as against £0.8 billion in 1938. Even a pessimist could hope that a cut to, say, £1.6 billion would prove feasible. This would have released some £3.6 billion gross for other purposes if

1. *National Income.* Cmd. 7371, p. 71, Table 21.
2. This optimism was fostered by ministerial statements that productivity had risen faster during the war than in the inter-war period, statements that were to mislead a number of careful observers (cf., e.g., Beveridge, *Full Employment in a Free Society*, Appendix C, p. 370).
3. E. g., L. Rostas, *Economic Journal*, 1943, p. 43; and T. Barna, *Bulletin of the Oxford Institute of Statistics*, July 1946. It is characteristic of the attitude of a broad segment of British opinion that either these facts were denied or claims advanced that the smaller output per head was more "economic" for Britain.

no change in the volume of national output had occurred. The net gain for civilian purposes would have been considerably smaller; for, apart from Lend-Lease, Britain in 1944 incurred £0.5 billion disinvestment at home and £0.66 abroad. But even so, *ceteris paribus*, some £2.7 billion were at the disposal of the Government for the increase in consumption and net investment, or over 30 per cent of the 1938 national income (the average price level having by then risen by some 50 per cent as compared with 1938).[4]

Even more important could have been (especially in the longer run) the contribution due to increased efficiency. To a large extent, improvement depended on capital investment and therefore could not at first be counted upon. But a great deal could have been accomplished by methods which did not absorb scarce resources, manpower and imports; the concentration of production on the most efficient plants, combined with double shift working; the standardization of products; the reorganization of the layout and working methods (including labor management), the rationalization of distribution, could be expected to yield handsome returns.[5] The spread of oligopoly and cartellization had so distorted British production that a major readjustment was necessary. At the same time the restrictive practices of trade unions and obsolete wage-systems which hindered efficiency also had to be altered. In many key industries, among which coal is notable, their unfavorable effect on production is considerable. Altogether, a coherent and determined effort at industrial reorganization might, even in the short run, have increased physical output by between 10 to 20 per cent. This represented another £500 to £1,000 million at 1945 prices.

Had the Government succeeded in mobilizing this output-capacity, the British problem could by now, in the main, have been solved. It is quite clear, however, that this is far from being the case, despite the generous help received from the United States — which moreover consented to waive those onerous conditions which first threatened to reduce its effectiveness.

Having regard to the extremely adverse conditions which Britain faced immediately after the war, her record of productive effort is more than creditable. The return to peace-time production demanded a huge re-deployment of manpower and considerable changes in equipment. Even a well-provided economy would have

4. D. Champernowne, "National Income," *Bulletin of the Oxford Institute of Statistics*, May 1946.
5. The most important firm in chocolate manufacturing estimated the increase in productivity as a result of the war-time introduction of some of these measures at well over a quarter.

taken some time for a full peace effort. And the British economy
had no commodity reserves; it was an "empty economy" at the
*beginning* (and not as some suggest at the *end*) of the process of
reconversion. The initial difficulties in the way of increased output
were proportionately greater and the time needed to overcome them
longer. As facts became available, charges repeatedly and vehe-
mently made that planning caused a fall in production (charges
mainly based on impressions derived from the fuel crisis which, though
disastrous, lasted only a few weeks) were disproved:

TABLE I

BRITISH INDUSTRIAL PRODUCTION

| | London,Cambridge[1] Economic Service 1946 = 100 | Central Statistical Office 1946 = 100 | | European Commission 1946 = 100[3] |
|---|---|---|---|---|
| | | 1935 weights | 1946 weights | |
| 1935............ | 100[2] | 88 | 95 | |
| 1936.......... | 110 | 97 | 104 | |
| 1937.......... | 117 | 104 | 111 | |
| 1938.......... | 109 | 98 | 104 | 95.7 |
| 1946.......... | 100 | 100 | 100 | 100[4] |
| 1947.......... | 109 | ... | 108 | 110.5 |
| 1948 (1 qtr).... | 120 | ... | 119.6 | .... |
| (2 qtr).... | 120.6 | ... | 121.6 | ....: |
| (3 qtr).... | 113.3̄ | ... | 113 | .... |
| (October). | 124 | ... | 127 | .... |
| (November) | 126 | ... | 129 | .... |

[1] Excludes finished munitions.
[2] The weights used were the geometric mean between 1935 and 1946 weights, the difference
being some 5 per cent in favor of the index with 1946 weights.
[3] Original base 1938.
[4] Last quarter extrapolated.

After the first war — much less exhausting and devastating in
its effects — British manufacturing output did not recover pre-war
levels until 1929 and then only for a year.[6]

The output of agriculture is some six per cent above pre-war.[7]

There can be little doubt that a very different impression was
created both at home and abroad by the general tenor of "expert"
British economic criticism of the Government's policy.

This relatively rapid redeployment and expansion of industry
was made possible by continued full employment (not experienced

6. League of Nations, *Industrialization and Foreign Trade*, 1945. Annex A.
p. 140. The somewhat earlier recovery of national real income must be attributed
to the change in the terms of trade in favor of Britain.

7. U. N. O. *European Economic Commission Report*, p. 11. Relative levels
comparable to the present were achieved only after 15 years in 1936.

between the two wars) and the absence of strikes. Unemployment, which in the whole of the inter-war period only twice dipped below 10 per cent (and then very slightly and for a few months) was reduced to less than three per cent. The loss of working days as a result of industrial disputes was 2.2 million days in 1946 and 2.4 million in 1947. In the first post-war period it had been 36.3 millions in 1919, 28.9 millions in 1920, 82.3 millions in 1921, and 19.7 millions in 1922. Continental countries which have decontrolled their economy show similarity to the first post-war pattern in Britain, especially in the number and duration of strikes. The increase in productive power of Britain as a result of avoiding them (apart from their adverse influence on general productivity) amounts to 1–2 per cent of output.

Nevertheless, a much greater effort would have been needed in order to restore the balance of the British economy, and this effort might have been made. The reasons for this relative failure are partly in the field of economic policy and even there they were only partially within the power of Britain to remedy.

(1) In the *international* field two reasons were important; the military and other external commitments of the Government, and the unfavorable turn in the terms of trade.

(*a*) The first of these not merely inflicted a loss of foreign exchange, but resulted in a massive absorption of manpower. In the middle of 1946 still just over two million men were in the armed forces as against less than half a million in 1939. To this has to be added some 0.3 million industrial staff employed by the services and an administrative staff of almost 180,000 as against 165,000 and 52,000 in 1939. Another 715,000 were working for the services in industry in 1946. Even by the middle of 1947 these numbers only diminished to 1.3 million, 277,000, 146,000, and 450,000, respectively. The unsettled conditions of the world no doubt demanded a higher force than in pre-war years. But it is arguable that immediately after the war the danger of a new outbreak was less, and the longer run strength of the country depended on a sound industrial base. On the average the tardiness of demobilization cost the country in the two years some 1.5 million man years, or some five per cent of national output. Relative to the United States — especially in terms of investment capacity — this represented a multiple drain on resources.

In 1946 the direct burden of the Government commitments abroad was £564 million or £194 million larger than the deficit in the balance of payments. In 1947 this fell to £374 million, while the deficit rose to £630 million or £256 million more than Govern-

ment expenditure. The indirect burden in terms of lost production might be estimated as at least £700 million, in these two years. The total burden was therefore well over the whole of the loan from the United States and Canada. No doubt some of this expenditure could not be avoided, but it would not be over-optimistic to claim that at least a large part could have been used to diminish our foreign deficit. This would have enabled Britain to use the United States loan to improve her capital equipment and replenish her stocks of raw materials and manufactures and thus accelerate the expansion of production. The cumulative effects of this can hardly be overestimated, for a small increase in total production means an appreciable narrowing of the gap.[8]

(b) The second factor which has retarded progress towards international balance was the worsening of the terms of trade. Until 1944 these were practically unchanged as compared with pre-war. Since then Britain experienced a deterioration of some 12 per cent, mainly as a result of the decontrol in the United States and the subsequent open inflation in the world markets. Had prices been unchanged as compared with pre-war the adverse balance of visible trade in 1946 would have been transformed into a surplus of almost £100 million and the deficit in 1947 decreased from some £440 million to less than £150 million.

Although this movement undoubtedly explains some of Britain's difficulties, it should not be attributed to temporary circumstances alone, and thus justify complacency. As it will be argued below, it does not as yet reflect the adverse turn due to the war in the *basic*

8. A further important cause of the loss of foreign exchange was the export of capital and repayment of war-debt. The latter has been especially sharply attacked and exceedingly high figures — up to £450 million — have been bandied about in the critical months after the suspension of convertibility. As a matter of fact, net repayment of sterling balances in the two years 1946 and 1947 amounted to only £100 million, a very small amount if we consider that the balances amount to some £4,000 million, and that most of them belong to poor countries in strategically vital areas. The suggestion that these balances could be wholly defaulted upon and the gold reserve (partly accumulated with the help of those countries who own them) exclusively used to maintain a (higher) British standard, seems unreasonable and certainly incompatible with either the restoration of London as an international monetary center or the possibility of further borrowing in this exceedingly cheap manner. Somewhat more questionable is the *legal* export of capital to the sterling area amounting to more than £300 million in the same period. Some of this was necessary for colonial reconstruction, but much of it was a "legalized" capital flight. And the relative laxity of the exchange control has made possible a substantive *illegal* flight of capital (mainly through understating export and overstating import values and through the sale of assets at less than full value). The control has now been tightened. Many loopholes were the direct consequence of the restoration of convertibility which made an adequate check of balances impossible.

relations of manufactures to primary products. It is still rather more favorable as yet than it was in the boom years of 1937 and 1929; and as the prices of primary products are notoriously more volatile than those of manufactures, the adverse shift could have been expected in conditions of world-wide prosperity. Even if the exceptional shortages and high prices due to the European crop failure in 1947 will be eased and with it the short run position of Britain improved, in the long run severe pressure must be expected on prices of manufactures as Europe and Japan revive.

The Government has failed — except for the so-called groundnut schemes and some contracts of the Ministry of Food (e.g., the Canadian wheat purchase) — to use bulk-purchase methods to help in solving this problem. Merely to centralize purchasing will remain a somewhat risky game in Britain in view of the fact that the selling countries might also adopt this expedient and they are in a better bargaining position under present circumstances to extort high prices. What was needed was the positive use of long-term guarantees to provide greater security to the producers by offering long-term contracts, and thus obtain favorable prices and stimulate an increase of the total supplies of primary products.[9] Most of the contracts, however, were for a year only and they covered mostly commodities already exported by the same areas prior to the war. Moreover, the Government agencies responsible for bulk purchasing were recruited from "existing trade channels" and worked through them. The checks on the policy of these "mixed" agencies, implied in the normal competitive pressure, disappeared. The result was, in many instances — especially food — the worst of both worlds.[1]

(2) The main *internal* causes for the failure to expand production fast enough to meet the external drain were partly connected with the manpower policy followed, partly with the inability to press for industrial reorganization. Progress in both fields was much impeded by the traditional resistance against reform of both employers' associations and trade unions, which was the direct consequence of the disastrous depression.

(a) Working population shrank considerably (apart from its partial unproductive absorption discussed above), from 21.5 millions in 1945 to 19.9 millions in March, 1948. The working week also decreased from a standard 48 hours towards 40–44 hours.[2] The two

9. The discontinuation of the long-term bulk purchase of nonferrous metals and other supplies has resulted in prodigious losses for Britain.
1. Cf., for a more detailed discussion, my article on the British Balance of Payments, *Bulletin of the Oxford Institute of Statistics*, Vol. 9, No. 7, p. 223.
2. Cf. *Annual Abstract of Statistics*, No. 84, 1935–1946.

together account for over ten per cent decline in potential maximum productive effort of the interwar period.

Part of this loss was inevitable. During the war married women and old people entered production, and their partial retirement was unavoidable. Relative wages of women have not improved — another sign of the inadequacy, from a national point of view, of collective bargaining — and taxation acted as a strong disincentive. The rate of retirement was greater than it need have, or ought to have, been. The decline in hours worked, which certainly had been excessive during the war, was also to be expected. Unfortunately, the push for a sharply reduced working week, and not merely for a reduction of the excessive overtime, was far greater than was either expected or compatible with a quick recovery. This is another instance where the strong bargaining position of the trade unions was not deflected by timely action to the benefit of the whole community.

(b) Apart from this serious loss of potential productive capacity, the Government consented to a rapid abolition of controls over manpower. Not merely was industrial conscription ended — a comprehensible and desirable step — but all measures enabling an *indirect* influence on the distribution of manpower, e.g., the provision that workers could only be engaged through official labor exchanges, and the restrictions on employers, were also withdrawn. The failure, in the name of non-interference with "free collective bargaining," to establish relative wages in accordance with the requirements; and the failure to reform obsolete wage-systems, which militate against increased effort; combined with the relaxation of physical controls over production and raw materials, retarded production. The monetary tension which tended to suck productive factors towards the "luxury" industries, which were not, or not as strictly, controlled as "essentials," thus caused a maldistribution of the labor force. Essential industries remained undermanned and — especially in coal and consequently in steel — acute bottlenecks originated, constricting the effort of the whole community.[3] Added to the lifting of controls under the last Churchill Cabinet and continued by the Labor Government in 1945–46, was the anxiety of the latter to avoid even transitional unemployment. Instead of enforcing as quickly as possible the once-for-all adjustment in the distribution of manpower necessitated by the reconversion to a peace-economy and more especially by the need to pay for imports by exports, the existence of the "pent-

3. Immigration was not sufficiently early and adequately tackled. The unions in some important instances at first refused to admit foreign workers for fear of renewed unemployment.

up" demand was permitted to stabilize employment in unessential but more congenial occupations.[4]  This need not have involved much or lengthy unemployment.  The creation of vested interest in the "new" (gadget) industries made subsequent action much more difficult.

(c) In contrast to the handling of the problem of obtaining a suitable *relative* wage level, the Government was successful in preventing a runaway increase in *absolute* wages.  The rise in wage rates from 150 in mid-1945 (September 1939 = 100) to 170 at the end of 1947 was certainly not cumulative and to some extent represented the readjustment due to the cessation of the war-time overtime.  In this success both subsidies and price-control, the abolition of which was demanded by the critics, played a vital part.  It has to be admitted, however, that an even earlier stand or at least an educational campaign against pressing for general wage rises would have been salutary.

(d) The post-war decontrol had other unfavorable effects.  Certain consumption-goods industries had been rationalized during the war by closing down some units and concentrating production on the more efficient plants or on factories whose manpower could not easily be used for the war-effort.[5]  The Coalition Government, not taking into account the problem of the balance of payments, promised to end these concentration schemes after the end of hostilities.  This promise was not reconsidered despite the continued need to use manpower as fully and as efficiently as possible.  In consequence an already inadequate manpower was spread over a greatly increased number of factories hardly any one of which could be manned fully.  Overhead costs rose and output per man declined.  In cotton spinning, e.g., one of the most important bottleneck industries, this decline was over ten per cent.[6]

4. The stimulus to the provision of electric appliances by reducing the purchase tax charged on them and providing scarce raw materials at a time when the insufficiency of generating plant was obvious, was one of the more glaring instances.  The decontrol of material allocation in the textile trade was responsible for the clothing muddle in 1948 when stocks accumulated.  In consequence, the home consumption of textiles was permitted to increase despite the fact that textiles remain one of the most remunerative exports.  Cf., below.

5. Unfortunately most of the concentration schemes were not effected by forming temporary holding companies as advocated by those who foresaw the probable post-war crisis, but by inter-firm agreements which did not safeguard the goodwill of the firms which had to close down.  Consequently the hardships involved by continuing it would have been much greater.  But even in the soft-drinks industry where no such hardship was involved, deconcentration was permitted.

6. Characteristically the deconcentration of petrol distribution was announced on almost the same day on which petrol allocations for private motoring were suspended altogether because of the convertibility crisis.

(e) Price control was considerably relaxed over a wide field. This, as we have already mentioned, caused a diversion of effort towards the non-controlled, unessential sector where profits rose. Where price-control was retained, it was based largely on a cost-plus basis or on the cost of the least efficient firms. This destroyed incentive for greater efficiency. Moreover, in many important fields, government intervention took the form of "voluntary gentlemen's agreements." All this strengthened the hand of trade associations; and the collusive administration of controls between government and industry further increased the oligopolistic and monopolistic features (already very ominous before the war) of the British industrial system.

Scarce resources were allocated with reference to the consumption in a base period — generally 1938 or 1939 — which had no relevance whatever to the present position and even less to the solution of the problems facing Britain. The entry of new firms was prevented or made more difficult. The rigidity and inefficiency of the system thus further increased. The restoration of "consumers' choice" in these conditions in effect meant the restoration of the producer's domination and his success in creating, through product differentiation, "secure" markets, and increased the number of models and types produced at the cost of efficiency and competitiveness. The relaxation of controls, moreover, resulted in a reduction of stocks in 1946 with unfavorable consequences on productivity. The coal crisis was only the most dramatic of these.

The British problem cannot be judged on the basis of United States experience. The eager striving for greater labor saving so characteristic of the United States remains almost completely absent in Britain. The self-restraints in British industry cannot easily be comprehended elsewhere. If the statutory attack on monopoly was largely unsuccessful in America the rather feebler attempts in Britain have no chance of success. Only an energetic combination of price-controls based on the cost of efficient producers, of compulsory standardization and of concentration, aided by suitable raw material allocation policies based on efficiency, could enable Britain to overcome her industrial handicaps without a severe deflationary crisis. There is no time to rely on the increase of efficiency through an extensive investment program. In Britain the results which in the United States might be brought about by decontrol and the inherent dynamism of the competitive market system could only be enforced by price and production controls. But such control needs to be far more efficient and ruthless than it has been hitherto.

(*f*) Nor has the government made full use of its own legislative and social program to promote efficiency. So far as the program of nationalization — except perhaps steel — is concerned, it was politically pre-determined and economically not controversial. But it should have been used for far-reaching reorganization of labor relations and of wage-systems. This did not happen. It would, moreover, have been prudent to make more obvious provision to *ouvrir la carrière au talent*. As it was, the nomination of trade union officials to high executive positons merely increased the internal struggle between the officials and the more ambitious of the rank and file, who wished to succeed them and therefore tried to outbid them for popular favor. Instead of canalizing talent towards a career within an industry through technical education, it was diverted toward political activity. The inevitable re-employment of the members of the old technical staffs, whose relations to the men had not been smooth, added to the grievances. The employment of erstwhile civil servants who found it difficult to grasp the human and business problems involved did not help. The increasing distance between the center and the working-places — an inescapable result of the amalgamation of small units into nation-wide corporations — was not bridged by suitable direct contacts. The result was indifferent organization and unofficial strikes, which detracted from the otherwise impressive success of, e.g., the nationalization of coal mines.

Nor was the large program of internal development — especially housing and social services — used to develop the mass-production of standard goods. As we have seen, the production of "standard" (utility) goods was discontinued over a large field. But the Government did not even use its power to guarantee large markets to tempt manufacturers into industrial reorganization. Had, e.g., the refrigerator industry been guaranteed the sale of 400,000 units over a period of years it could have cut costs to less than a third of those now prevailing. The importance of such a move for the export drive cannot be too much stressed. Altogether, it is no exaggeration to conclude that the Government has missed the chance of increasing British production by at least another 15–20 per cent — however creditable its accomplishment may be in comparison with the results achieved after 1918 or with those in European countries in which "freedom" was restored.

(*g*) The combination of these factors made itself inevitably felt on exports where control was also considerably released. The priority of export orders prescribed was not backed by strict raw material allocations. Nor was the licensing of exports continued.

Consequently the spectacular recovery of exports from 45.8 per cent of their pre-war volume in 1945 to 99.3 per cent in 1946 did not continue in the first half of 1947. Indeed a relapse occurred.

VOLUME OF EXPORTS (1938 = 100)

| | |
|---|---|
| 1945 (year)................... | 45.8 |
| 1946 (year)................... | 99.3 |
| (4th quarter)............. | 111.8 |
| 1947 (1st quarter)............. | 100.6 |
| (2d quarter)............. | 101.9 |
| (3d quarter)............. | 113.9 |
| (4th quarter)............. | 118. |
| 1948 (1st quarter)............. | 125.6 |
| (2d quarter)............. | 134.3 |
| (3d quarter)............ | 138.2 |
| (4th quarter)............ | 146.3 |

Moreover, exports to "soft" currency countries, which were more profitable at the current rates of exchange, showed a substantially greater increase than those to areas from which essential food and raw-material imports had to be obtained. Devaluing sterling would not have helped under these circumstances (even if offsetting wage-increases could have been avoided and export prices increased to the maximum possible so as to avoid a violent worsening in the terms of trade), as the *relative* attractiveness of foreign markets would not have been influenced by this measure. More could have been achieved by refusing to sell except against "available" sterling (i.e., sterling balances held by foreign countries as a result of current transactions)[7] or gold. And the attraction of the home market had to be countered. Despite the fall in the productive potential and the continued drain on civilian manpower, consumption was permitted to increase from 1945 to 1946 by some ten per cent, and in 1947 increased further by another three per cent. It may not have recovered to the pre-war level[8] but the fall, even per head, is small indeed.[9]

7. In distinction to "old" pre-war or war-time balances.
8. The White Paper estimates a slight increase but it does not however take into account quality changes and probably understates the rise in prices. Cmd. 7371, p. 41.

PERSONAL EXPENDITURE AT 1938 PRICES (£ MILLION)

| 1938 | 1944 | 1945 | 1946 | 1947 |
|---|---|---|---|---|
| 4,288 | 3,706 | 3,921 | 4,292 | 4,424 |

9. Cf. G. D. N. Worswick, "A Fall in Consumption," *Bulletin of the Oxford Institute of Statistics* (June 1948), Vol. 10, No. 6. It is interesting to compare the facts now available with the wild statements on production and consumption by the critics of the Government made while neither production nor consumption indices were available. The fact that the high income tax promotes claims for expenses which do not appear in the national income figure leads to an under- rather than to an overstatement of the productive effort.

It is the non-availability of goods which command high consumers' surpluses, the deterioration in the stock of durable consumers' goods (including houses), and the greater social equality which had to be obtained by reducing the share of the more prosperous and therefore more articulate classes of the community, which has created the impression of grave impoverishment. The supply of durable goods, including houses, increased sharply, and there was a substantial expansion in other — non-essential — services and goods, — the obverse effect of the decontrol over production which has been discussed above. There can be no doubt that rationing, price-control, and subsidies to essentials created an artificial demand for non-essentials. But this does not mean that the measures taken in respect of the former were unjustified. It means that they should have been accompanied by controls necessary to maintain equilibrium.

The fact that shortages have led to the adoption of *fair shares* in the distribution of *necessities* does not eliminate the need for inequality of income to give incentive. Only the incentive is by common consent limited to (semi) *luxuries* so as not to involve hardship. It behooves an advanced community so to educate its members that discipline should be possible even without the whip of starvation. On the other hand, if such a system is established, the encroachment of unessentials on the total effort must be prevented by direct controls. The anxiety of the Government to avoid even short-term frictional unemployment was to some extent responsible for their failure to deal with this problem. The insistent demand for derationing and high consumption, backed by reputable experts, however, made the necessary "austerity" almost impossible.

The Government's record is thus better than the critics admit. The alternative policies advocated would, as in France and Italy, have led to social trouble and economic frustration. The Government's failures — in a position of unparalleled difficulties caused by the war — were due to its ceding to popular clamor (to a large extent supported by the majority of expert opinion within and without the Government) both in relaxing controls over production and in permitting so large an increase in consumption. Moreover, an inordinate and invaluable period was wasted until a central planning agency was established and properly staffed. Consequently there was an inexplicable delay before it was realized that the country's external commitments were out of balance with its economic strength, before action was taken to readjust them, and also, before steps were taken to increase and secure a balance in productive capacity.

IV. THE ELEMENTS OF A SOLUTION

The external commitments of the country, the worsening of the terms of trade, together with the internal decontrol and the consequent misdirection of production and undue increase in consumption, led to a rapid deterioration of the current balance of payments of Britain in 1947. The loss of reserves due to incomplete control and capital exports was even worse.

U. K. BALANCE OF PAYMENTS (MILLIONS) [1]

|  | Imports f.o.b. | Exports and re-exports f.o.b. | Balance on visible trade | Invisibles (net) | Total balance on current account | Total loss of reserves |
|---|---|---|---|---|---|---|
| 1946 |  |  |  |  |  |  |
| 1st half | 502 | 392 | − 110 | − 120 | − 230 | 80 |
| 2d half | 590 | 498 | − 92 | − 48 | − 140 | 146 |
| 1947 |  |  |  |  |  |  |
| 1st half | 725 | 511 | − 214 | − 121 | − 335 | 469 |
| 2d half | 815 | 591 | − 224 | − 71 | − 295 | 555 |
| 1948 |  |  |  |  |  |  |
| 1st half | 887 | 731 | − 156 | + 16 | − 140 | 254 |

[1] H. C. Deb., Vol. 456, No. 8 (23rd September, 1948), Columns 137 and 138. Figures for the second half are not yet available. The dollar deficiency was more than covered by E. R. P. By the end of the year an over-all balance was more or less reestablished, although the dollar deficit continues.

Despite accumulating signs of crisis (and, apparently, the willingness of the United States Government to waive the respective clauses of the Anglo-United States Financial Agreement) convertibility was introduced in July, 1947. It had to be suspended within a month. As the Administration had expressed complete assurance in its ability to fulfil the obligations which it had undertaken despite repeated warnings, confidence was severely shaken and a far-reaching reconsideration of policy took place.

It was agreed on all sides that measures had to be taken to bring effective demand and supply into a more balanced relation. The question was whether this was to be done by global measures restricting monetary demand, while freeing both it and supply of direct controls; or was the nominal money demand to be more severely controlled, its disturbing effects on the distribution of productive resources curbed, and steps taken to increase supply by direct intervention and stimulus. In this way the pent-up demand could also be eliminated or at least sterilized. In the end, a compromise was arrived at — mainly under the pressure of political forces — and a mixture of these policies was applied. In the foreign field the strengthening of direct controls prevailed — if for no other reason because

time (and reserves) were lacking to test the alternative. In the domestic field the "global" measures were more predominant. It will be shown that their consequences were not altogether favorable.

In the field of foreign economic policy the retreat towards bilateralism was the more precipitate as the confidence of the experts in the soundness of the experiment in non-discrimination and unilateral convertibility had prevented any preparations of an alternative policy, and no reserves remained. Imports from the Western Hemisphere were sharply cut. Despite a further rise of more than 10 per cent in foreign prices and increasing need for raw materials due to the expansion of output, the Government managed to keep imports from rising proportionately. Controls over exports were reintroduced and the system of priority allocations was considerably reinforced. This produced a spectacular increase in the volume of exports from roughly 100 per cent of 1938, in the first half of 1947, to almost 150 per cent in the middle of 1948, an advance the rapidity and extent of which is hardly paralleled by any country.[1] Despite misgivings voiced, Britain did not find much difficulty in concluding a considerable number of trade and payments agreements which brought its balance of trade more into equilibrium and prevented further losses of dollars and gold. Despite (or perhaps rather because of) the dropping of convertibility, she succeeded in limiting the drain on accumulated sterling balances and in limiting the legal and illegal flight of capital. Equally important was the cut in the gross cost of foreign commitments of the Government from £564 million in 1946 to £236 million (annual rate) in the first half of 1948, and in net cost from £291 million to £114 million. As the foreign income of the Government was temporary, the cut in gross cost was essential. Altogether the foreign balance improved from a deficit of £335 million in the first half of 1947 to only £140 million in the first half of 1948. This was all but within the limits of the Marshall Aid which had in the meantime been put into execution by the United States, thus effectively reversing the American foreign economic policy hitherto pursued. In the second half of 1948 a further considerable improvement took place.

The British domestic economic policy also underwent considerable, though less abrupt, changes. These can be divided into two groups: on the one hand the relaxation of controls was reversed and measures were taken to limit Government commitments to set free productive resources. On the other hand, and mainly under the

1. See table, p. 48. The figure for July, 1948 was 149. The target for June was 140; for December, 150. (*Economic Survey for 1948.* Cmd. 7347, p. 8.)

impact of the violent criticism against Government policy suggesting that the breakdown had been caused by inflation, steps were initiated to cut money demand. The "cheap money" policy was reversed, taxation was increased, subsidies reduced, and investment projects cut.

A closer analysis shows that the former steps had a far greater effect, and moreover that the latter had some unfavorable repercussions.

Elaborate calculations have been made[2] to demonstrate that a dramatic change in the British monetary background has taken place at the turn of the new financial year (April 1 in Britain), usually accompanied by eulogies of Sir Stafford Cripps' new policy. The Chancellor certainly deserves high praise for his determination to dispel that air of unreality which surrounded the discussion of both domestic and foreign economic policy, and to call attention to the grave long-run position of the country. But the turn of events cannot by any means be attributed solely or even mainly to the change in the monetary background. The autumn budget of 1947 increased prospective revenue by some £250 million. It was harshly criticized as insufficient. Yet taxation was not further increased. The changes in taxation decided upon in the 1948–49 budget were "inflationary" on any sensible definition: direct taxation on income was cut by £103.5 million while indirect taxation was increased by some £40 million only. The capital levy, true enough, is expected to bring in £105 million, but an overwhelming part will certainly be paid out of capital and will not diminish outlay. The change from a "true" deficit of £432 million in the calendar year 1947 to a "surplus" of £315 million in the financial year April 1948–March 1949, is certainly dramatic — almost £750 million, or little less than 10 per cent of the national income, at factor cost.[3] But, as we have seen,[4] from an economic point of view these calculations are rather too "schematic." It is not true, e.g., that the sale of surplus stores has no "disinflationary" effect in a position where pent-up demand is not restricted to the consumers' sphere. On the other hand, many "capital" repayments, e.g. war-damage payments, are "sterilized" by effective controls on capital expenditure; and income tax payments, though they fall due seasonally, are anticipated by the taxpayer in his decisions of spending. The change was much more continuous

---

2. E.g., R. C. Tress, "The Budget and the National Income," *London and Cambridge Economic Service,* Vol. XXVI, No. II (May, 1948).

3. Ibid., p. 43.

4. Cf. above, p. 35.

than these figures suggest and came about more through a decline of expenditure than through an increase in revenue.

The abandonment of "cheap" money resulted in an increase of the rate of interest from roughly 2½ to 3 per cent on medium Government stock. This small change certainly could not influence investment decisions. Even a much greater rise would hardly have had substantial effects as the liquid reserves of firms made them independent of the capital market. It is questionable, moreover, whether in Britain at present the usual market process could satisfactorily differentiate between different projects even if it could be made effective. The relative attractiveness of different projects could only be compared by that method if prices were free. But the freeing of prices would not merely raise them well above long-run equilibrium prices but would also start a violent inflationary spiral which would render all comparison impossible. However imperfect the control of capital investment, if handled with some common sense, it serves the purpose more satisfactorily. A rise in interest rates without decontrolling prices would not reflect "natural" relations but merely indicate differences in the price policy of the different controls. In any case, when risk is absent (as it largely is in a fully employed country) a small rise would be completely ineffective as profit margins are of a higher order of magnitude. A sharp rise, on the other hand, is impracticable for budgetary reasons.

It is sometimes suggested that "dearer" money, by diminishing stock exchange values, at least depressed consumption power. Even this is questionable. Those speculators who were active and spent freely when prices were increasing had liquidated their holdings. The trusts and institutional holders of securities will hardly have decreased their outlay (the increase in spendable *income* as against capital appreciation was the much-complained-of result of the cheap money policy) as a result of a depreciation of capital. Although there were strong arguments against a *continued fall* in interest rates, their *rise* must in the long run be harmful.

Even more questionable was the wisdom of the proposal to cut investment, put forward by a section of economic experts.

Investment in fixed capital (including housing) was only £1,200 million in 1946 as against £750 million in 1938. But the price of capital goods had more than doubled in the meantime. In 1947 gross investment rose to just under £2,000 million[5] and was to be

5. *National Income.* Cmd. 7371. In an earlier White Paper on Capital Investment the figures given were £1550 million for the middle of 1947 and £1600 million for the middle of 1948. These figures exclude changes in the value of stocks and only take account of certain kinds of investment in housing and

reduced by some 15 per cent. To those economists who claim that the *global* amount of investment was too ambitious, the answer is simple. The war-damage and obsolescence not made good was very heavy, amounting to probably not less than £3,000 million. The capital equipment continues to be exhausted by its fuller rate of utilization. The depreciation allowances even in 1938 amounted to some £450 million or some £1,000 million at present prices. The Government measure (which was a much watered form of the proposals of the orthodox economists) meant in fact a reduction of net investment to, at most, £750 million or, in terms of 1938 prices, less than £400 million, i.e., roughly equal to the insufficient investment of the slump year of 1938. This cut was administered at a time when the United States rate of investment was, on any calculation, at a record level and United States industrial productivity had sensibly increased above pre-war, which in its turn was a multiple of the British. The campaign which suggested that hardships could be avoided and both consumption and the balance of payments simultaneously restored by diminishing capital investment was not merely based on the crude statistical fallacy of counting the factors released doubly. It also had most regrettable psychological consequences by suggesting that the crisis was of a superficial character and easily curable without discomfort now or in the future. Nothing could have been more unfortunate.

There was, however, a more specific and less extreme criticism against the investment program. This attacked partly the costliness and elaborateness of the housing program. Moreover, it was suggested that the lack of central co-ordination, to which reference has already been made, resulted in an inflation of investment projects beyond the capacity to undertake them simultaneously, thus causing unnecessary congestion, bottlenecks, delay, and a decrease in productivity. The review of investment projects was rightly concentrated on limiting further schemes until those under execution were completed. This had an immediate and favorable influence. Actual cuts in investment, on the other hand, were fortunately not drastic. The suspension of new projects had, however, a discouraging effect on enterprise, the consequences of which were to show themselves subsequently. The standard of housing accommodation, however,

industry. The cuts were to have amounted to some £230–250 million. Cf. *Capital Investment in 1948.* Cmd. 7268; H. M. Stationery Office. Cf. also the admirable critique, "Cuts in Capital Investment," by F. A. Burchardt, in the *Bulletin of the Oxford Institute of Statistics,* Vol. 10, No. 1 (January, 1948). The cuts were concentrated on the planned increase in road vehicles (50 per cent) while plant and machinery were to suffer less (20 per cent).

which had been improved considerably and thus impinged on the quantity of housing which could be provided, was not reviewed and adjusted at all.[6]

The most important factor on the monetary side was, however, the stabilization of the pent-up demand, considerably increasing the effective level of net saving. At the moment of gravest apprehension, a reduction in the prospective gross investment in fixed capital[7] of the country to £1,550 million was aimed at. It was thought that £700 million of this could be covered by private means, of which undistributed profits were to have provided some £225 million. Taking into account the dissaving implied in the decrease of tax reserves, this would have necessitated personal, private saving of some £450 million, a high but not impossible achievement out of a private disposable income of almost £9,000 million. All these calculations have been upset by the increase in private saving, the expansion of production which had not been taken into account, and the consequent improvement in the budget and in company finance, as well as by the grant of Marshall Aid. By the summer of 1948 it became likely that the surplus and capital maintenance of public authorities would be nearer to £425 million and £145 million respectively; Marshall Aid and other foreign disinvestment, £300–400 million; business depreciation allowances, £800 million. The total sum available is, therefore, £1,670–1,770 million, excluding private saving. Plowed back profits will not be less than £350 million and might be higher. Thus £2,000 million of gross investment could be financed *without* any private saving. The Government plans foresaw only £1,550 million of investment including all investment financed out of private personal saving. Private saving will hardly be less than £250 million (it was £450 million in 1947). Thus the flow of savings and the investment capacity of the country was gravely underestimated.

The combination of the agitation against austerity, against controls, and against investment had prevented or retarded the preparation of investment plans to make immediate use, for an increase in productive investment, of such unemployment as might arise, since the advisers of the Government still thought in terms of runaway inflation. Nor had the redeployment of labor been organized with sufficient care and thoroughness. The Government was therefore forced to retreat, both on the cuts in consumption and in investment;

6. Cf. above.
7. It can be argued, of course, that pre-war housing standards were too low and that it would be unfortunate to saddle the country with housing which would soon have to be improved. On the other hand, the change in the size of families had increased the demand for smaller houses.

especially since the suspicion of the trade unions had been aroused, partly by the rise in prices (due at first mainly to increased taxes and the cut in subsidies, but subsequently to the rise in foreign prices), and partly by the propaganda against "over-full-employment," which was directed mainly against their strong bargaining position. Subsidies were increased from an annual rate of £390 to £470 million, thus decreasing the budget surplus. The clothing ration was increased and certain items (shoes, furniture) altogether de-rationed.[8] The home supply of other goods was also permitted to increase. This meant a loss of potential exports. The housing program was restored to almost its original volume, and the maintenance of 200,000 houses under construction is envisaged instead of only the contemplated 120,000. Much more serious was the decision to increase the limit of house repairs permissible without license from £10 to £100 per annum per house — a measure which might increase the volume of non-essential and unproductive investment by some £150–200 million, a drain not easy to support.

Thus the "disinflationary" program had the paradoxical result of diverting goods, which should have been exported, to the home market and of absorbing in house repairs resources which should have been used to increase *productive* investment. Of course it could be (and has been) argued that the Government should not have yielded to the political pressure resulting from unemployment and loss of domestic trade and should have maintained pressure on demand until all readjustments had been made and the "essential" industries been manned. This argument is reinforced by a plea that decontrol of domestic prices, the cut in subsidies to the price of essentials, and devaluation would have further reinforced the effectiveness of the deflationary policy.

The case for devaluation is completely fallacious as long as no resistance is encountered in export markets, since it would turn the terms of trade against Britain. Britain imports food and raw materials which have well-organized markets and world prices and she exports manufactures which are subject to market imperfections. For this reason, a worsening in the terms of trade might be serious enough to offset any stimulus to the volume of exports. Imports are, in any case, held well below their "free" level and their quantum would not be affected by devaluation. The rest of the argument is even weaker. In order to obtain a margin in favor of export prices

8. The accumulation of stocks, which forced this measure, was also the consequence of the decontrol over the allocation of raw materials, which led to a misdirection of supplies. Cf. above, Part III.

as large as is maintained at present by controls (the index of average export values is at 251 (1938 = 100) as against a domestic wholesale price index of 218,[9] mass unemployment of a severe kind would have to be created.[1] This fact is hardly ever faced by those who argue that controls are a "blunt" weapon, incapable of working efficiently.[2]

Similar considerations must also lead to rejection of the domestic aspects of this policy. The unemployment required to accomplish the necessary shift of manpower without controls and without a conscious wage policy would be so high as to lead to a breakdown of the policy of co-operation between the unions and the Government. A "slight disinflation" would certainly not have sufficed. The advice to raise prices by decontrol and cessation of subsidies would, far from leading to deflation and a shift into export industries (because people would have less money to spend on non-essentials), lead to violent wage demands and increase the inflationary impact. Moreover, under those conditions the holders of liquid reserves would no longer hold back: an explosive cumulative process might ensue.

The rejection of this policy by the Government was amply justified by the fact that the maintenance of social peace and the expansion of production contributed far more to the better balance between supply and demand than any other factor, a fact often slurred over in discussion of the British problem.

Industrial output increased in 1947 by eight per cent over 1946. But the improvement was most striking in the last quarter of the year, which ran some 20 per cent above the 1946 average. In the first half of 1948 this advance was maintained. The consequent increase in supplies can be put at some £1,000 million at wholesale

9. This comparison is not strictly justified because the export index is based on part of exports only. But it is indicative of the position.

1. An even greater degree of unemployment would be required to balance foreign trade without quantitative controls.

2. So much also for the fashionable view that bilateral agreements do not increase foreign trade under certain circumstances. British capacity to import was safeguarded and increased by the hurriedly negotiated (and therefore somewhat too restrictive) trade agreements after the breakdown of the unilateral convertibility, which in its turn resulted in a competitive restriction on British exports in the futile effort to obtain dollars from her. It must be admitted, however, that a number of complications might arise (not the least because of the need to spread overhead cost over a sizeable home market) which have as yet not been faced. The end of the sellers' market would aggravate these problems because decisions would have to be reached as to how far to permit sales at home of products unsaleable abroad (at any rate for a period needed to readjust production). But for all this, the strong pressure exercised by priority raw-material allocations for exports and by the control of domestic prices is proven by the extraordinary expansion of the volume of exports in a period of full employment at home, an expansion which belies the doubts expressed by the "disinflationists."

prices and the increase in the supplies for civilian purposes at £1,300 million, or almost double even the *nominal* cut in monetary demand in consequence of budget policy (the *effective* cut was much less[3]). The diminution of Government expenditure abroad was matched by a decrease in domestic expenditure. Including only that on goods and services (i.e., excluding transfer payments) this decreased from £4,495 million in 1945 to £2,473 million and £2,168 million, respectively, in 1946 and 1947, with a further expected drop of some £100 million in 1948. Hence there has been a substantial release of manpower for productive work of well on to 1.5 millions, or almost 10 per cent, since the middle of 1946, and 0.4 millions since the middle of 1947, which became effective only after a certain lag. The production available for export and domestic civilian purposes thus increased even faster than total output.

These achievements were partly the automatic consequence of the process of demobilization and reconversion. But to a large extent they were due to the gradually accelerating reversal of the decontrol initiated during the Coalition and Conservative Governments in 1945 and continued by the Labor Administration. It was at last recognized that the maintenance of full employment necessitated planned intervention if inflation was to be avoided; and that Britain's international balance could not be regained without undue and unnecessary sacrifice (and thus the risk of social strife paralyzing production, as happened in some Continental European countries), if direct controls were not strengthened. As it was the pressure of events, rather than conscious policy, which brought this change about, it came rather belatedly, and after precious time had been lost.

The control over employment was restored. This does not mean industrial conscription. It merely means that instead of the power of the purse combined with the whip of unemployment, an intelligent and discriminatory influence is exerted on employers' hiring policy. The charge that slavery is about to be established in Britain has been exposed to the ridicule of facts. At the same time, the pressure for higher wages which had been growing in 1947 has been checked by voluntary co-operation on the basis of a Government statement.[4] The most important cause of an inflationary spiral was thus stopped. The whole outlook now hinges on the success of this policy. Should the voluntary system fail, some sort of statutory power will have to

3. Cf. above, p. 56.

4. The increase in wages was to some extent a result of reconversion to peacetime production. As wages in the "shrinking" industries could not be lowered, those in the "undermanned" trades had to be raised. But the process threatened to assume a cumulative character.

be used.   Unfortunately, less attention was paid to establishing a relative wage structure which would help, rather than hinder, the redeployment of labor, and little success has been achieved in reforming the outdated wage system so as to maximize incentive.

A further increase in prices would obviously end all hopes of stability.   The diminution of the subsidies to the price of essential goods, combined with the rise in world prices, has already put a grave strain on labor relations.   Any further cut, which seems still to be favored by some, would certainly unleash violent wage demands and inflation.[5]   For the same reason the extension of control over the price of non-essential goods is an essential factor of recovery. True enough, the Government is only groping towards an effective policy on prices.   Control is still based either on cost-plus or on the cost of the less productive firms.   It is, moreover, in many cases still only "voluntary," and therefore precarious, often exploitative and often inefficient.   But a beginning has been made and it is to be hoped that a steady pressure will be exerted on the oligopolistic and monopolistically competitive price structure which had such fatal consequences on British efficiency between the two wars.

Domestically the policy will be tested by its success in increasing production, and this can only come from increased productivity.[6] Since the first half of this year not much progress has been made. While all agree on the need for higher efficiency, neither the Government nor the two sides of industry seem willing to undertake the necessary reforms with vigor.   Apart from intensified investment, which is of longer-run influence only (and which depends on whether the Government is strong enough to resist demands for increased consumption), efficiency could in the short run be increased by technical changes, such as better layout of factories and the standardization and concentration of production.   Some experiments have been made with good results (e.g., in cotton spinning), but they remain on a small scale.   Nor has there been much progress made in establishing industrial development councils — the organization of which was permitted by an Act already two years on the Statute Book — which could undertake this task.   The resistance encountered by the proposal to set up an Anglo-United States joint advisory council to explore the problem shows a disturbing spirit of complacency, and a

5. This has been recognized by the Government.   The food subsidies, which according to the budget speech were to have been limited to £400 million, have lately been increased to £470 million.   Cf. my paper on "The Abolition of Food Subsidies," *Bulletin of the Oxford Institute of Statistics,* November, 1948.
6. On the methods of increasing productivity, cf. Part III.

reluctance to face facts.  Nor was there much accomplished by way of reforming obsolete systems of pay which reduce incentive and thus obstruct progress.

The worsening of the international situation and the consequent drain on manpower and production is a renewed threat to recovery, especially in relation to the foreign balance.  That aspect of the British problem is now dominated by United States policy and the Marshall Plan.  An increase in physical production which would restore the balance in foreign relations by 1952 is well within the limits of British efforts.  But the conditions which might be attached to United States aid might easily endanger the balance of payments.[7] The time has not yet come to discuss the probable consequences of an almost simultaneous cessation of aid to European countries, with the entry into force of the non-discrimination clauses of the I. T. O. Charter (not to speak of the obligation to maintain convertibility which has not been formally abandoned), as the formation of a Western European payments and trading area might profoundly alter prospects.  Although the basic condition of recovery, social peace, has been preserved (despite the demands to pursue a policy leading to unemployment and a fall in real wages) and substantial progress made towards a reintegration of Britain into a more balanced world economy, the future of British recovery still hangs in the balance.

## V. Conclusions

These considerations suggest that criticisms of the policy of the British Government, coupled with demands for decontrol, disinflation, and devaluation are not substantiated by a closer analysis of the economic developments in the country, and of the problems faced by it.  These criticisms were based mainly on hasty impressions created — in the absence of reliable and up-to-date statistics — by the fuel crisis of 1947 and on theoretical misconceptions of the nature of the inflationary pressure.  Moreover, they ignore the social implications and thus the indirect economic consequences of the (monetary) measures advocated.  The reconsideration of the salient determinants of the British economic position also suggests the directions in which a solution could be found.

(1) There was and remains general agreement that the relation-

7. For a more detailed discussion, cf. my articles, "Exchange Depreciation and Economic Readjustment," *Review of Economics and Statistics* (November, 1948), and "Discrimination" and "Intra-European Clearing," *Bulletin of the Oxford Institute of Statistics*, Vol. 10 (1948), Nos. 7, 8, and 9.

ship between *effective* monetary demand and supply has to be brought into balance, and thus the international balance of payments restored. The dispute on British policy centers on the *method* of readjustment only and not on the aims to be attained. The choice is between an attempt to return to a "free" market system on the one hand, and a strengthening of direct controls in order to prevent a misdirection of resources, on the other. We have seen that in the short run, decontrol, deflation and devaluation could not bring about effective monetary balance, not merely because of the much greater relative volume of war-time pent-up demand in Britain as compared with the United States, but also because stocks of commodities are still extremely low and cannot be increased because of the balance of payments crisis. This disparity would inevitably start a cumulative inflationary spiral once controls were removed. The scarcity of stocks would accentuate the inflationary momentum by encouraging speculation. By maintaining controls and speeding the increase of productivity on the one hand, and by budgeting for a surplus on the other, the pent-up demand might be eliminated in the longer run. This demand will also tend to press less on the markets as supplies begin to increase, and as hopes rise of being able to purchase in the future on better terms. The volatility of the pent-up demand enjoins careful planning in handling deflation, as otherwise unemployment might be caused suddenly if liquidity preference rose, and the man-power set free might not be absorbed in undermanned industries and in investment. The consequent political pressure might result in an undue increase in consumption (including durable goods). Conclusions drawn from an analysis of the British or European position, however, are not necessarily applicable elsewhere.

(2) The inflationary spiral in Britain (and in many other parts of Europe) would almost certainly become explosive in the absence of controls if unemployment were not extremely heavy, for otherwise wage demands would be pressed home and the bargaining power of the trade unions would be too strong. It is fallacious to assume that the level of unemployment which is needed without controls (a) to keep wages stable, (b) to stimulate exports, and (c) to balance international payments, is identical and equal to the level of unemployment which is inevitable as a result of people passing from job to job. This is far from being the case. The tighter the trade union organization, the larger the "industrial reserve army" needed in an uncontrolled economy to maintain monetary stability. "Disinflation" cannot be "successful" in the sense of permitting *decontrol* without creating *unemployment.* The social consequence of a deliberate policy aiming

at maintaining a large industrial reserve army such as the "disinfla-
tionists" advocate, would be incalculable.  It is noteworthy that all
European countries which attempted to re-establish a "free" market
economy have menacing Communist movements.  In contrast, those
countries (mainly Scandinavia, Holland and the United Kingdom)
which have controlled systems have escaped social strife.  Once the
working class is organized and the realization spreads (through the
popular acceptance of the Keynesian view) that the level of employ-
ment is not an unalterable consequence of the interplay of natural
forces, the basis for the functioning of "free" markets is undermined
(except perhaps in countries which possess an agricultural reserve
population or in which technical progress is high enough to set free
workers at the rate required to maintain relative stability without
distress e. g. the United States).

(3) The success of such "disinflation" as there has been in Britain
seems due to accelerated demobilization and a reduction of foreign
military commitments.  Thus productive capacity increased substan-
tially.  Moreover, controls were reimposed and the distribution of
manpower adjusted.  The continued wastage through absorption
in non-essential production was effectively limited.  The inflationary
pressure was relieved by the voluntary agreement of labor not to press
for higher wages.  This agreement was based on the maintenance of
relatively stable prices and the continuation of the policy of full
employment and social reform.  It is unlikely that the consent of
labor could have been obtained on other terms.

Indeed, when the pressure of pent-up demand abated and when
unemployment threatened, the Government was forced by the conse-
quent political pressure to reverse its deflationary policy.  This was
the case both in consumers' goods (clothing, furniture, house-repairs)
and in producers' goods (factory building).  In consequence the
opportunity to increase productive capacity and exports was lost.
More careful planning might have avoided this misdirection of effort.
The Government has not solved the once-for-all problem of the large-
scale shift of manpower which arose as a result of reconversion to
peace production.  The relative maldistribution of manpower will
therefore have to be relieved slowly.  But no such problem is likely
to arise again (unless war breaks out).  The lesser day-to-day read-
justment required by shifts in demand can be handled much more
easily.

(4) If the suggestion that decontrol, combined with monetary
measures aimed at reducing real effective demand, could solve the
British short-run problem, ignores the social, and thus the indirect

economic, consequences of the policy proposed, it is perhaps even more to be criticized from a longer-run point of view.

(a) The abolition of key controls (as contrasted with their gradual relaxation and the reduction and simplification of control) would seem to require the permanent maintenance of a large reserve of unemployed. A co-ordinated wage and price policy, coupled with measures to steer workers into occupations where they are most needed, and with a control on the hiring policy of employers, will reduce this wastage through unemployment.[8]

(b) But apart from the elimination of instability, it is fallacious to argue that decontrol in Britain would restore a perfectly competitive price structure and accelerate capital accumulation and economic progress. The British economic problem in this respect differs basically from that of the United States. All evidence points to the conclusion that decontrol would result in the continuation of trends towards monopolistic restrictionism which are responsible for the present inadequacy of British capital equipment and for inefficiency. Anti-monopoly legislation could not restore a workable system, for it cannot deal with such tacit agreements, which have been prevalent in Britain.[9] Moreover, the need for increased efficiency would seem to depend on a technical reorganization of industry involving large-scale amalgamations, concentration of production, and standardization of products. Such large-scale changes are hardly practicable (and would certainly not be undertaken) without Government intervention. This should be both positive, e.g., stimulus through guarantees of sales, provision of capital (either finance, or building and leasing of factories); and negative, e.g., pressure on excessive profit margins through maximum price limits. The threat of control and nationalization has already had very salutary effects in, e.g., the steel and automobile industries. Moreover such reforms could not be put into effect without the closest co-operation of the trade unions. This would not be forthcoming in Britain without some Government safeguards for the workers. It is not an exaggeration to conclude that *controls are needed* in Britain to *enforce efficiency*. Investment must also be speeded up. Voluntary savings are unlikely to be forthcoming in sufficient volume to sustain the required volume of

8. Cf. (2) above.
9. This would also affect the problems of stability, as fewer workers would be freed by technical progress ("technological unemployment"); and, as (unlike the United States) we have no manpower reserves in agriculture, the maintenance of monetary equilibrium (difficult enough even in the expanding economy of the United States) would be entirely dependent on a substantial degree of unemployment. Cf. (1) and (2) above.

investment without a considerable increase in the inequality of income distribution. Such a reversal of social policy would, however, be incompatible with the aims of the Government and with social peace and monetary stability: wage demands would certainly follow on any step which increased profits. While some tax-reform encouraging the plowing back of profits (especially by young and smaller firms) might be carried through, the increase in savings will, in the main, have to come through the maintenance of a substantial budget surplus. This should not be used for "disinflation," as some have proposed, but for an increase of the investment both by the publicly owned sector and — through finance companies — also of private firms. Positive State intervention will have an important role to play in reconciling a more equal distribution of income with the need for quicker economic progress.

(5) Britain — and no doubt certain other European countries — must look for her salvation to an improvement of the existing system of controls and to their co-ordination on the basis of a coherent plan. This was acknowledged with great foresight by the United States Administration of the Marshall Aid Program when it demanded the submission of detailed four-year plans from the recipient countries. In the last year considerable progress has been made in both directions. The Government has also begun to harmonize financial and physical controls. Much still remains to be done. Price controls and raw-material allocations must be thoroughly reformed to stimulate efficiency. Obsolete wage systems also remain to be altered in order to improve the incentive to effort while maintaining monetary stability. There is grave danger that the fear of resentment will drive the Government further towards a "tripartite" administration of controls, or in other words towards a collusion with well organized pressure groups on both sides of industry not dissimilar to the "corporative" state. This would hinder an increase of productivity from its present unsatisfactory level, since reforms enforcing its increase would no doubt run counter to vested interests. Britain needs more information and far more publicity to improve both the work of planning and that of intelligent criticism. Much can be learned from the United States in this respect.

(6) The long-run prospects for the British balance of payments cannot be reviewed in this context.[1] It will depend chiefly on factors

---

1. Cf. my articles "Discrimination" and "Intra-European Clearing," *Bulletin of the Oxford Institute of Statistics*, Vol. 10, Nos. 7 and 9, and "Exchange Depreciation and Economic Readjustment," *Review of Economics and Statistics* (November 1948), esp. Part IV.

which are beyond the control of the British Government, especially on the position and policy of the United States. It is obvious, however, that quantitative controls of imports and direct stimulus to exports through domestic price control and priority raw material allocations, must be maintained. The Marshall Plan vouchsafes a certain breathing space. It will have to be used — in conjunction with the internal re-organization — to establish a strong and efficient economic system in Western Europe and in those overseas countries with which Western Europe has special political ties and which are in the main complementary to its productive structure. Such close co-operation must thus take the form of preferential arrangements which permit the rationalization of British and Western European industry and the expansion of its markets overseas. While the Marshall Plan insists on the need for such — discriminating — use of European supplies for European reconstruction, there is no indication as yet that the United States realizes that these arrangements cannot come to substantial fruition before the expected end of the aid program (1952), but must be continued for a considerably longer transition period. Only in this manner can the ruptured multilateral system of world trade, based on the export surplus of Europe to tropical and subtropical areas and the import surplus from North America, be restored. It is also questionable whether the flimsy structure of reconstruction could survive a cessation of United States grants, without other non-U. S. areas (at least non-European) being enabled by loans or U. S. import surpluses to buy European products against dollars. Criticism of the British Government in this respect has become widespread and it is suggested that it hinders the establishment of Western European unity. This criticism is not understood in London and arouses violent resentment. The fact is that the British ministers and experts (apart from the legitimate doubts about formal customs or federal unions) are still laboring under the concepts underlying the abandoned policy of non-discrimination. It is not that they wish to hinder progress towards closer co-operation, but that they lack the burning conviction needed to establish a co-ordinated plan for Western Europe, and to carry it through by sheer moral leadership against mounting technical obstacles due to the open inflationary disintegration in many European countries.

(7) The task before the British Government is formidable. There are no short cuts to its solution. But however distant the goal, the real achievements of Britain must not be ignored. The talk about falling production and muddle is refuted by the facts, which show that output is 25 per cent above pre-war and still rising. The

real danger to continued British recovery must be considered to lie mainly in the threat of renewed expansion of military expenditure (especially direct expenditure abroad) beyond the country's economic capacity. It is to be hoped that the Government will not commit the errors of 1945–46 once more, and that it will approach the United States betimes to consider the incidence of the cost of what is a common foreign policy. Recovery might also be checked by premature decontrol and by an undue increase of consumption permitted for reasons of internal politics. The hostile comments of British experts give a false picture abroad of Britain's effort and success, and should be refuted if Anglo-United States co-operation is to thrive.

### POSTSCRIPT

This essay was written early in 1948 and brought up to date in September. In the months which have passed since, the triumph of the policy here adumbrated continued. Manufacturing output has increased one-third above pre-war, a record never achieved in the twenty years of peace after the first World War. An *overall* balance in international payments is within grasp. The advocates of return to free-market economics and deflation have, once more, proved completely wrong. Nevertheless popular and "expert" pressure, and the attraction of the "liberal" viewpoint to the highest levels of the civil service has, as in 1946–47, led to a renewed partial reversal of the policy of balanced control. This might well weaken the force of recovery and even result in a relapse.

Thus on the domestic side the program of the Government for the next four years submitted to E.C.A. envisages too small an increase in productivity, i.e., 15 per cent in four years, hardly more than that of the inter-war period which left Britain well behind its chief competitors. This is due in part to the failure to provide for an increasing rate, in terms of the national income, of investment, and in part to the apparent scepticism of the Government of being able to take effective steps to reorganize British industry.

Some features of the Government plans in the international field seem also questionable. (1) The recent change in the U. S. attitude opens the way to the re-establishment of a multilateral world economic system, such as was never possible under Bretton Woods; and the utmost use should be made of it. The plans published imply, however, that Britain has not yet sufficiently reorientated her policy of retreat towards bilateralism, and more particularly, that it is cutting its dollar deficit in far too mechanical a way, instead of using a fuller U. S. aid both to increase its productive investment and to

promote European recovery. (2) The success of the multilateral approach implies a much greater degree of economic cooperation than obtains at present. It is regrettable that the British Government has not taken the initiative in strengthening the O.E.E.C. so that it can fulfil the task of a really effective international planning agency. (3) In the longer run multilateralism can be maintained only if some of the rules of the I.T.O. are modified and the length of the period of transition extended. Moreover cooperation of the United States in solving the dollar problem is essential, including toleration of transitional discrimination against her and one-sided tariff concessions. It also implies effective help in promoting progress in under-developed areas.

It is to be hoped that the coming months will see a thorough revision of the British four-year plan and the growing cooperation of the British and the United States Governments in extending the scope of international planning to include an effective solution of the above problems.

T. BALOGH.

BALLIOL COLLEGE, UNIVERSITY OF OXFORD

# THE SECULAR TREND IN MONETARY VELOCITY

## SUMMARY

Introduction: Change in monetary velocity as a measure of change in the demand for money, 68. — Types of money and their respective velocities, 69. — Velocity data pertinent to analysis of business fluctuations, 71. — Controversial issues regarding monetary velocity, 74. — The secular trend in circuit velocity of money, 75. — The trend in velocity and growth in time deposits, 81. — Seasonality of production and the rate of use of money, 84. — Factors in the declining trend in the rate of use of money, 86. — Interrelations between changes in the quantity and velocity of money, 90.

The primary function of money is to serve as a means of payment for goods and services which are purchased; the secondary function is to serve as a store of value which can readily be used as, or converted into, means of payment whenever the holder, in an emergency situation or some other circumstance, wishes to make purchases of larger value than his current income or cash receipts allow. The desire of people for a stock of money for these two purposes was described by eighteenth and nineteenth century economists as a need for a medium of exchange and a need for a store of value. In recent economic literature the same concepts are frequently referred to as a demand for money for transaction purposes and a demand for money to hold. Both groups of economists recognize that the strength of the demand for money for these two purposes depends upon the habits and customs of the population.

Development of a measure of change in the demand, or need, for money for these two purposes has been a neglected phase of the application of quantitative concepts and analysis to economic phenomena. In the eighteenth and nineteenth centuries this was due primarily to the paucity of factual data; in recent times it appears to have been due to a failure to realize the relation between a dynamic or growing economy and the need for money, and to the form in which changes in the demand for money are expressed in the available statistical data.

In a dynamic economy where population is growing and improvements in techniques of production are continuously being made, full use of economic resources (to use a basic concept of orthodox economic theory) or full employment (to use a contemporary phrase with essentially the same meaning) involves a continuous increase in the quantity of money of the nation; except for the possibility of varia-

tion in the habits of use of money by individuals and enterprises, or of a decline in the level of prices. If stability in the price level is assumed to be desirable, in order to avoid the inequities of inflation on the one hand and the tendency for a general decline in prices to produce a business depression on the other, full use of resources or full employment requires a continuous growth in the quantity of money comparable with the rate of growth in production, except for adjustment to changes in habits of use of money. The argument which occasionally appears that the money supply is ample or excessive because it is greater than it has ever been before is an obvious reflection of a lack of attention to the dynamic or growing character of the economy of the United States.

The failure of contemporary economists to realize the form in which changes in the demand for money are expressed in statistical data appears to be due to an erroneous assumption that changes in the quantity of money reflect changes in the demand for money — that is, that changes in the quantity of money are dominated by factors on the demand side. This assumption carries with it the assumption that in practice the decisions which simultaneously increase the volume of bank assets and the money supply (currency and bank deposits) are made by customers of the banks. Scrutiny of the conditions under which loans and investments are acquired or relinquished by the banking system — that is, of the dominant forces influencing the total amount of bank loans and investments — shows that this assumption is incorrect. The quantity of money is dominated by factors on the supply side; that is, by the decisions of bank officials respecting their loans and investments, and by the conditions established by law and central bank operations under which bank officials make those decisions.

Changes in habits of use of money — that is, changes in the need or desire for money other than those directly associated with growth in population and in productive capacity — are not reflected statistically in the quantity of money held by business and individuals; but in ratios of that quantity to the various types of payments which they make for the output of the economy, or in connection with the processes of production and marketing. The reciprocals of such ratios may be used as measures of the velocity or rate of use of money.

### Types of Use of Money and Their Respective Velocities

Payments, or transfers of money from one person or firm to another, may be grouped into three major classes: (1) those associ-

ated with the production and marketing of the current products of the economy; (2) those associated with the transfer of ownership of various types of property, whether land, durable goods, or titles to such goods, from one person or firm to another; (3) those constituting movements of money other than transactions associated with current production or transfers of ownership of property. A recent writer has designated payments in the first class, "transactions in the main money circuit," and those in the other classes, "technical transactions."[1]

Crude estimates only are provided by available data, either for the total volume of payments or for the volume in each of the three major groups. The closest approximation to the total volume of payments currently available is an estimate of the amount of debits to bank accounts other than interbank.[2] Such debits do not include transfers of currency from one person to another, but in effect they cover payments made in cash to the extent of cash withdrawals from deposit accounts. They probably provide a fair measure of the degree of change from one year to the next in the total volume of payments.

A rough estimate of the total amount of the first of the three major groups of payments — those associated directly with the production and marketing of the current output of the economy — can be prepared by adding together estimates of expenditures for the final products of the economy, sales of wholesalers, sales of manufacturers, sales of primary producers, and total income payments. Such a summation is not published currently. A comparison of annual estimates of this total with the amount of bank debits, for the period 1919–1947, indicates that payments directly associated with the production and marketing of current products have varied

1. Morris A. Copeland, "Tracing Money Flows Through the United States Economy," *American Economic Review*, XXXVII: Proceedings (May 1947), p. 40. Mr. Copeland classifies technical transactions into three principal types: (1) money-changer transactions, which occur when a transactor makes a payment to himself, as when he cashes a check, buys foreign exchange, or transfers a balance from one bank to another; (2) agency transactions in which one party acts as a disbursing or collection agency for another; and (3) financial turnover transactions of two sorts; namely, portfolio liquidations and investments to the extent that such transactions offset each other, and repayments of indebtedness and new borrowing to the extent that these transactions offset each other. In the classification of payments which I have made all three types of technical transactions, except payments relating to portfolio liquidations and investments (i.e., transfers of property) would fall into my third group, "movements of money other than transactions associated with current production or transfers of ownership of property."
2. The estimates are made on the basis of debits reported by banks in a large number of centers to the Board of Governors of the Federal Reserve System,

in aggregate amount quite differently from total payments in the economy.[3]  The former, as estimated, declined from 40 per cent of the latter in 1923 to 26 per cent in 1929, and then rose to 41 per cent in 1935 and to 55 per cent in 1942.  From 1923 to 1929 estimated payments directly associated with current production and marketing rose 19 per cent but total debits rose by 86 per cent.  From 1929 to 1933 estimated payments associated with current output were reduced by 50 per cent, but total debits dropped by 66 per cent.  From 1933 to 1941 estimated payments associated with current output rose by 139 per cent, while total debits rose by 73 per cent.

Information regarding the comparative amount of payments associated with transfers of property and those representing other movements of money (the second and third categories, respectively) is so scanty that it appears impracticable to attempt estimates of the amount of either of these two major types of payments.  However, the assumption is reasonable that the great variation since 1923 in the ratio of estimated payments associated with current output to total debits is due primarily to variation in the amount of payments associated with transfers of property.  The period, 1923–1929, when debits rose far more rapidly than payments associated with current output, was the well-known period of rapid increase in sales of real estate and securities.

VELOCITY DATA AND THE ANALYSIS OF BUSINESS FLUCTUATIONS

The phrase "business fluctuations" may be applied to all fluctuations in business including transfers of property.  However, in common usage the term refers to variations in the current output of the economy and to the fluctuations in employment associated with those variations.  The monetary velocities, therefore, which are most pertinent with respect to business fluctuations are those which relate to the use of money for payments associated with the output of goods and services.  Such velocities may be of several types, of which the most comprehensive is the ratio of all payments associated with current output to the total quantity of money in the nation.  This ratio, however, as computed from available data, has a low degree of reliability because of the crudity of the estimates of the amount of payments.

Another type of ratio which represents the velocity or rate of use

3.  The estimates are given in Table I.  Mr. Copeland's estimates of "transactions in the main money circuit" for the years 1936–1942 lead him to the same conclusion.  He states:  ". . . the aggregate of technical transactions sometimes exceeds the aggregate of transactions in the main money circuit and . . . the relationship between the two is somewhat eccentric" (op. cit., p. 43).

TABLE I

RELATION OF CIRCUIT VELOCITY TO OTHER MEASURES OF MONETARY VELOCITY, ANNUALLY, 1919–1947

| Year | Estimated payments (billions of dollars) | | Ratio of total payments associated with current activity to: | | Payments per $1 of money (average total quantity)[3] | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| | Debits to deposit accounts except interbank (commercial banks)[1] | Total payments associated with current activity[2] | Debits to deposit accounts | Value of final products sold | Debits to deposit accounts[4] | Total payments associated with current activities[5] | Payments for final products[6] |
| 1919 | 663.0 | 274.0 | .41 | 3.78 | $18.38 | $7.59 | $2.01 |
| 1920 | 721.0 | 296.6 | .41 | 3.90 | 18.20 | 7.49 | 1.92 |
| 1921 | 591.0 | 216.0 | .37 | 3.30 | 15.68 | 5.73 | 1.73 |
| 1922 | 643.0 | 226.6 | .35 | 3.36 | 16.52 | 5.82 | 1.73 |
| 1923 | 685.0 | 276.2 | .40 | 3.57 | 16.21 | 6.54 | 1.83 |
| 1924 | 716.0 | 273.8 | .38 | 3.46 | 16.05 | 6.14 | 1.77 |
| 1925 | 820.0 | 293.6 | .36 | 3.48 | 17.11 | 6.13 | 1.76 |
| 1926 | 872.0 | 308.6 | .35 | 3.45 | 17.42 | 6.16 | 1.79 |
| 1927 | 952.0 | 300.7 | .32 | 3.38 | 18.27 | 5.77 | 1.71 |
| 1928 | 1,114.0 | 309.7 | .28 | 3.41 | 20.55 | 5.71 | 1.68 |
| 1929 | 1,276.0 | 327.7 | .26 | 3.47 | 23.45 | 6.02 | 1.74 |
| 1930 | 931.0 | 279.7 | .30 | 3.33 | 17.49 | 5.26 | 1.58 |
| 1931 | 685.0 | 221.4 | .32 | 3.10 | 13.26 | 4.29 | 1.38 |
| 1932 | 471.0 | 168.9 | .36 | 3.01 | 10.32 | 4.70 | 1.23 |
| 1933 | 437.0 | 165.4 | .38 | 3.11 | 10.35 | 3.92 | 1.26 |
| 1934 | 491.0 | 192.9 | .39 | 3.13 | 10.58 | 4.16 | 1.33 |
| 1935 | 547.0 | 222.7 | .41 | 3.32 | 10.69 | 4.35 | 1.31 |

| Year | | | | | | |
|------|------|-------|-----|------|-------|------|
| 1936 | 628.0 | 258.0 | .41 | 3.35 | 11.33 | 4.65 | 1.39 |
| 1937 | 650.0 | 288.9 | .44 | 3.48 | 11.34 | 5.04 | 1.45 |
| 1938 | 566.0 | 259.9 | .46 | 3.26 | 9.75 | 4.48 | 1.37 |
| 1939 | 592.0 | 279.2 | .47 | 3.31 | 9.44 | 4.45 | 1.34 |
| 1940 | 627.0 | 309.7 | .49 | 3.37 | 9.26 | 4.58 | 1.36 |
| 1941 | 756.0 | 394.8 | .52 | 3.45 | 9.99 | 5.22 | 1.51 |
| 1942 | 854.0 | 472.3 | .55 | 3.13 | 9.95 | 5.50 | 1.76 |
| 1943 | 1,055.0 | 558.0 | .53 | 2.99 | 9.33 | 4.94 | 1.65 |
| 1944 | 1,187.0 | 604.1 | .51 | 2.97 | 8.77 | 4.46 | 1.51 |
| 1945 | 1,297.0 | 613.1 | .47 | 3.02 | 8.14 | 3.85 | 1.28 |
| 1946 | 1,398.0 | 673.1 | .48 | 3.63 | 8.09 | 3.90 | 1.10 |
| 1947 | 1,498.0 | 791.7 | .53 | 3.73 | 8.89 | 4.70 | 1.26 |

[1] 1919–1941, *Banking and Monetary Statistics*, p. 254; 1942–1947, extrapolated on basis of debits in reporting centers, *Federal Reserve Bulletin*, July 1947, p. 868 and May 1948, p. 532.

[2] Sum of the following items: (1) payments by business and individuals for final products, author's estimate in "Quantity and Frequency of Use of Money," *Journal of Political Economy*, LIV (October 1946), pp. 438–41, subsequently revised and extended; (2) estimated government expenditures other than purchases of goods and services, National Income Supplement to *Survey of Current Business*, p. 23; (3) wholesale sales, 1929–1944 *Statistical Abstract of the United States, 1944–45*, p. 863, with prior years extrapolated on basis of ratio to value of manufactures, and subsequent years extrapolated on basis of ratio to sales of service and limited wholesalers, *Survey of Current Business*, February 1948, back cover; (4) value of manufactures in odd-numbered years, 1919–1939, *Statistical Abstract of the United States, 1944–45*, p. 795, with intervening years interpolated on basis of Shaw's estimated value of finished commodities at producers' current prices, *Finished Commodities since 1879*, p. 6, and subsequent years from Department of Commerce estimates (*Survey of Current Business*, February 1948, back cover) or interpolated on basis of Department of Commerce index of value of manufacturers' shipments, *Survey of Current Business*, July 1946, p. 23, and later issues; (5) farm cash marketings, *Statistical Abstract of the United States, 1944–45*, p. 648, and *Survey of Current Business*, February 1946, p. 32, and later issues; (6) value of mineral products, 1919–1943, *Statistical Abstract of the United States, 1944–45*, p. 750, with subsequent years extrapolated on basis of indexes of mineral production and prices of selected mineral products; and (7) personal income, excluding income in kind not involving payments, from data in National Income Supplement to *Survey of Current Business*.

The data used in these estimates contain some duplication, but omit some payments associated with current activity. For the years 1936–1942 the amounts given here range from 90 to 95 per cent of Copeland's preliminary estimate of "total transactions in the main money circuit," loc. cit., p. 40.

[3] Average total quantity of money from estimates in "Quantity and Frequency of Use of Money," loc. cit., subsequently revised and extended.

[4] While debits to deposit accounts do not directly reflect payments made in currency, they include the portion of such payments represented by checks cashed at banks.

[5] For 1936–1942, corresponding amounts computed from Copeland's "total transactions in the main money circuit" are as follows: 1936, $5.15; 1937, $5.34; 1938, $4.83; 1939, $4.81; 1940, $4.91; 1941, $5.52; 1942, $6.13.

[6] From estimates in "Quantity and Frequency of Use of Money," loc. cit., subsequently revised and extended.

of money associated with current economic activity is a ratio of expenditures for final products to the average total quantity of deposits and currency, excluding interbank obligations. This kind of ratio, or circuit velocity of money, is the simplest type of statistical series available for indicating the change from time to time in habits of use of money associated with current production and distribution of goods and services. Two varieties of this kind of ratio have been prepared by the author and published elsewhere.[4] These series are advantageous in studies of business fluctuation not only because of their availability and simplicity, but also because they focus attention on the final products of each year's operation of the economy and conform to the theoretical framework of the equation of exchange applied to the value of those final products.

### CONTROVERSIAL ISSUES REGARDING MONETARY VELOCITY

Comments on monetary velocity in contemporary economic literature indicate the presence of conflicting opinions as to the following questions respecting the rate of use of money:

1. Is the apparent downward trend since World War I in the circuit velocity of money a true secular trend? If so, is it of recent origin or of long standing?

2. Does this secular trend, if it exists, relate only to the total quantity of money, defined to include time deposits; or does it apply also to demand deposits and currency, generally assumed to represent the active portion of the money supply? That is, does the computed downward trend in circuit velocity of money reflect only the growth of time deposits?

3. If this trend is due to factors other than growth of time deposits, what is the nature of those forces?

4. What are the relations between deviations from trend in the rate of use of money and deviations from trend in the quantity of money? Are such deviations generally synchronous or sequential in time; and are they generally offsetting or do they operate in the same direction with respect to their relation to the ups and downs of business?

Fortunately, the available statistical data, while not as abundant for earlier periods as in recent years, are sufficient to provide reasonably conclusive answers to these questions.

4. "Quantity and Frequency of Use of Money in the United States," *Journal of Political Economy,* LIX (October 1946), pp. 436–50. One of these ratios is based on the value of all final products sold; the other on payments for such products by business and individuals.

# THE SECULAR TREND IN MONETARY VELOCITY    75

## THE SECULAR TREND IN CIRCUIT VELOCITY OF MONEY

A downward trend has existed since the close of World War I in the circuit velocity or rate of use of money, when measured either by the ratio of all expenditures for the final products of the economy to the total quantity of money (including time deposits), or by the ratio of such expenditures by individuals and business enterprises to the quantity of money which they hold. Nevertheless, as in the case of other secular trends, it is impossible to obtain from those data a clear answer to the question of whether this trend is the result of fairly continuous changes in the habits of use of money over a long period of time, or is the result of special circumstances operating during a part or all of the period since the close of World War I. Professor Albert G. Hart, who has examined the data for this period and similar data for a few years prior to 1919, has concluded: ". . . it appears clear that the 'trend' is a fiction." Professor Alvin Hansen, on the other hand, refers to an historical tendency, over a period of more than a century, to hold an increasing amount of money in relation to income.[5] The difference between these two interpretations of the data is of considerable practical importance for current economic policy. If the present low ratios of expenditures for final products to the quantity of money reflect primarily the results of the abnormal world-wide

### FIGURE I
RATIO OF NATIONAL INCOME TO THE QUANTITY OF MONEY 1799–1939
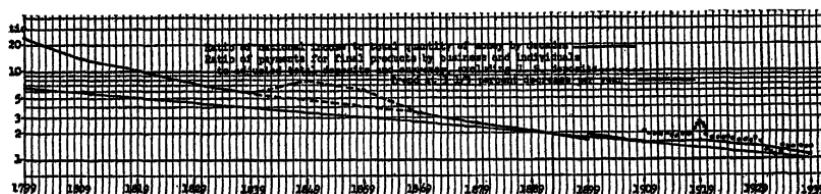
situation during and since the nineteen thirties, there is danger of a more serious postwar price inflation than if they reflect primarily a declining trend in the rate of use of money.

In Table II estimates are given of the national income, the quantity of money (deposits plus currency excluding that owned by banks), and the ratio between them by decades from 1799 to 1939. The ratios are also shown on Figure I, together with annual ratios of

5. Albert G. Hart, "Postwar Effects to be Expected from Wartime Liquid Accumulations," *American Economic Review*, XXXV: Proceedings (May 1945), p. 343; Alvin H. Hansen, *Economic Policy and Full Employment* (1947), pp. 216 and 326.

## TABLE II

TOTAL DEPOSITS AND CURRENCY, EXCLUDING INTERBANK OBLIGATIONS, AND REALIZED NATIONAL INCOME, BY DECADES, 1799–1939

(Amounts in millions of dollars)

| Year[1] | Money in circulation (including cash in banks)[2] | Bank cash (specie)[3] | Bank deposits, excluding interbank[4] | Total deposits and currency, excluding bank cash and interbank deposits | Realized national income[5] | Ratio of national income to total deposits and currency |
|---|---|---|---|---|---|---|
| 1799 | 27 | 9 | 10 | 28 | 677 | 24.2 |
| 1809 | 55 | 15 | 25 | 65 | 915 | 14.1 |
| 1819 | 67 | 20 | 40 | 87 | 876 | 10.1 |
| 1829 | 87 | 22 | 65 | 130 | 975 | 7.5 |
| 1839 | 220 | 49 | 111 | 282 | 1,631 | 5.8 |
| 1849 | 233 | 44 | 127 | 316[6] | 2,420 | 7.7[6] |
| 1859 | 439 | 105 | 388 | 722[6] | 4,311 | 6.0[6] |
| 1869 | 741 | 149 | 1,437 | 2,029 | 6,827 | 3.4 |
| 1879 | 819 | 190 | 1,994 | 2,623 | 7,227 | 2.8 |
| 1889 | 1,380 | 488 | 4,232 | 5,124 | 10,701 | 2.1 |
| 1899 | 1,904 | 703 | 7,474 | 8,675 | 15,364 | 1.77 |
| | | | | Total deposits adjusted and currency outside banks June 30[7] | | |
| 1899 | ... | ... | ... | 8,036 | 15,364 | 1.91 |
| 1909 | ... | ... | ... | 15,794 | 26,456 | 1.68 |
| 1919 | ... | ... | ... | 35,605 | 62,945 | 1.77 |
| 1929 | ... | ... | ... | 55,171 | 79,498 | 1.44 |
| 1939 | ... | ... | ... | 60,943 | 69,815 | 1.15 |

Some of the data, for the period 1799–1829, are from reports or estimates for a year or two subsequent to the year specified.

¹ 1799–1859, *Annual Report of the Comptroller of the Currency*, 1920, Vol. 2, p. 49, with the amounts shown here for 1799, 1809, 1819, and 1829 taken from Treasury estimates for the respective subsequent years.

² 1799, assumed to be one-half of total specie, as in 1811; 1809, 1819, and 1829, Gallatin's estimates for 1811, 1820, and 1829, respectively, *Annual Report of the Comptroller of the Currency*, 1876, p. xl; 1839, 1849, and 1859, *Annual Report of the Comptroller of the Currency*, 1876, p. xlv; 1869–1899, A. Platt Andrew, *Statistics for the United States* (National Monetary Commission, Senate Document No. 570, 61st Congress, 2d Session) p. 33. Figures for 1869 are for national banks only; those for 1879 do not cover private banks and probably not all State banks; while those for 1889 include cash items for banks other than national banks.

³ Figures for 1799–1859 relate to deposits in State banks, savings banks, and the Bank of the United States, and do not include amounts reported as due to banks. They are based on data published in the *Annual Report of the Comptroller of the Currency*, 1876, pp. v–xlv and lxxxiii–cxxii. The original data for 1799–1829 consist chiefly of Gallatin's estimates for State banks in 1811, 1820, and 1829, his statements for the first Bank of the United States in 1809 and 1811, annual statements for banks in Massachusetts for 1803 and subsequent years, and estimates by E. W. Keyes of deposits in savings banks in 1820, and statements for the second Bank of the United States in 1830. For 1799 and 1809, the estimates are based largely on data for circulation and the amount of deposits relative to circulation for later years. The original data for 1839, 1849, and 1859 consist of reports by State officials collected by the Secretary of the Treasury and reported to the House of Representatives and published as House documents. See footnote 6 for a comment on the adequacy of these data.

⁴ Figures for 1860–1899 are estimates of total deposits in all commercial and mutual savings banks, based partly on data collected by the Comptroller of the Currency and partly on tax returns submitted to the Bureau of Internal Revenue, minus estimates of interbank deposits. The estimates of total deposits are from the *Annual Report of the Federal Deposit Insurance Corporation*, 1943, pp. 103–108, and 112–118; those for interbank deposits from the *Annual Report of the Comptroller of the Currency*, 1920, Vol. 1, p. 261.

⁵ Estimates for 1799–1929 are from Robert F. Martin, *National Income in the United States*, 1799–1938 (1939), pp. 6–7; estimate for 1939 is from National Income Supplement to *Survey of Current Business*, p. 19, adjusted for comparability with Martin's estimates. For this adjustment the following items are omitted: corporate profits tax liability, undistributed corporate profits, inventory valuation adjustment, food and clothing furnished government employees, and services furnished without payment by financial intermediaries. The total thus obtained is then reduced by 2.3 per cent on the basis of the two estimates of realized income for 1929.

⁶ The figures for both deposits and currency for 1849 and 1859 are too small, and the ratios therefore too high. Subsequent to the expiration of the charter of the second Bank of the United States in 1836, agitation, and in some States legislation, against chartered banks led to a rapid development of private banks; and the adoption of the "free banking" system in a large number of the States resulted in establishment of many State banks not required to submit reports of assets and liabilities to State officials. These influences probably do not appreciably affect the figures collected by the Secretary of the Treasury for 1839, but they evidently do affect them for later years. The total deposits of State banks in 1849 is given as $91.2 million, compared with $90.2 million for 1839; but for 21 States for which figures are given for both years, the deposits in 1849 were 35 per cent larger than those in 1839. The figures for State banks for 1859 are more complete than those for 1849, but the lack of data for private banks makes an estimate of the degree of underreporting impracticable.

⁷ *Banking and Monetary Statistics* (1943), pp. 34–35.

TABLE III

CIRCUIT VELOCITY OF MONEY INCLUDING AND EXCLUDING TIME DEPOSITS, ANNUALLY, 1909–1947

| Year | Payments for final products by business and individuals per $1 of:[1] | | | Relative growth of time deposits[5] | |
|---|---|---|---|---|---|
| | Average deposits and currency held by business and individuals[2] | Adjusted total deposits and currency (excluding U.S. Government deposits)[3] | Adjusted demand deposits and currency (excluding U.S. Government deposits)[4] | Time deposits per $100 adjusted total deposits and currency (excluding U.S. deposits) | Time deposits in commercial banks per $100 total time and adjusted demand deposits (excluding U.S. deposits) |
| 1909 | .... | 2.01 | 3.34 | $39.84 | $28.98 |
| 1910 | .... | 1.96 | 3.33 | 41.03 | 30.58 |
| 1911 | .... | 1.86 | 3.18 | 41.42 | 31.18 |
| 1912 | .... | 1.95 | 3.35 | 41.95 | 32.02 |
| 1913 | .... | 1.95 | 3.43 | 43.17 | 33.51 |
| 1914 | .... | 1.80 | 3.10 | 41.82 | 30.58 |
| 1915 | .... | 1.74 | 3.14 | 44.74 | 34.88 |
| 1916 | .... | 1.84 | 3.21 | 42.68 | 33.71 |
| 1917 | .... | 1.93 | 3.34 | 42.25 | 34.27 |
| 1918 | .... | 2.05 | 3.37 | 39.24 | 32.68 |
| 1919 | 2.12 | 2.07 | 3.39 | 38.75 | 32.59 |
| 1920 | 2.07 | 1.99 | 3.32 | 40.03 | 34.88 |
| 1921 | 1.84 | 1.78 | 3.21 | 44.37 | 38.95 |
| 1922 | 1.82 | 1.74 | 3.15 | 44.91 | 39.11 |
| 1923 | 1.94 | 1.83 | 3.42 | 46.49 | 41.36 |
| 1924 | 1.88 | 1.77 | 3.38 | 47.63 | 42.43 |
| 1925 | 1.86 | 1.75 | 3.36 | 47.93 | 42.61 |

| Year | | | | | |
|---|---|---|---|---|---|
| 1926 | 1.91 | 1.78 | 3.49 | 48.99 | 43.71 |
| 1927 | 1.84 | 1.71 | 3.46 | 50.62 | 45.25 |
| 1928 | 1.82 | 1.67 | 3.47 | 51.89 | 46.44 |
| 1929 | 1.91 | 1.72 | 3.59 | 51.98 | 46.18 |
| 1930 | 1.71 | 1.54 | 3.29 | 53.19 | 47.10 |
| 1931 | 1.42 | 1.32 | 2.91 | 54.65 | 47.90 |
| 1932 | 1.25 | 1.17 | 2.59 | 54.70 | 47.00 |
| 1933 | 1.28 | 1.21 | 2.58 | 53.25 | 43.75 |
| 1934 | 1.40 | 1.32 | 2.71 | 51.41 | 41.38 |
| 1935 | 1.41 | 1.31 | 2.55 | 48.60 | 38.52 |
| 1936 | 1.47 | 1.35 | 2.51 | 46.14 | 36.47 |
| 1937 | 1.59 | 1.47 | 2.72 | 45.88 | 36.69 |
| 1938 | 1.50 | 1.38 | 2.58 | 46.50 | 37.49 |
| 1939 | 1.47 | 1.34 | 2.41 | 44.26 | 35.28 |
| 1940 | 1.47 | 1.34 | 2.29 | 41.33 | 32.56 |
| 1941 | 1.55 | 1.43 | 2.31 | 37.94 | 29.95 |
| 1942 | 1.50 | 1.40 | 2.11 | 33.79 | 26.99 |
| 1943 | 1.43 | 1.34 | 1.93 | 29.83 | 24.16 |
| 1944 | 1.37 | 1.30 | 1.86 | 30.25 | 25.70 |
| 1945 | 1.24 | 1.19 | 1.74 | 31.69 | 27.87 |
| 1946 | 1.32 | 1.21 | 1.79 | 32.69 | 28.82 |
| 1947 | 1.44 | 1.34 | 2.01 | 33.48 | 29.40 |

Footnotes to table on next page.

1 Payments for final products by business and individuals, 1910–1947, are estimates from "Quantity and Frequency of Use of Money," *Journal of Political Economy*, LIV (October 1946), pp. 438–41, subsequently revised and extended. Such payments for 1909–1918, used in these computations, are as follows, in billions of dollars: 1909, 31.6; 1910, 33.2; 1011, 33.0; 1912, 36.6; 1913, 37.7; 1914, 36.0; 1915, 35.8; 1916, 44.5; 1917, 52.7; and 1918, 61.2. These estimates are derived from the following sources: Lough's estimate of consumers' spendings, excluding imputed items, in 1909 and 1914 (*High Level Consumption*, pp. 236–45) with interpolations for 1910–1913 and 1915–1918 on basis of Department of Commerce estimate of such expenditures (*Basic Facts on Employment and Production*, p. 12); Department of Commerce estimate of value of construction for 1915–1918 (*Survey of Current Business*, July 1947, p. 28) with extrapolations for 1909–1914 on basis of Newman's index of construction (*The Building Industry and Business Cycles*, pp. 65–67); value of selected manufactured products representing producers' durable equipment, 1909, 1914, and 1919, with interpolations for intervening years on basis of Shaw's estimated value of such goods at producers prices, *Finished Commodities since 1879*, p.6; government receipts extrapolated from 1919 on basis of National Industrial Conference Board estimates for fiscal years, *The Economic Almanac for 1948–43*, pp. 410 and 411.

2 Estimates from "Quantity and Frequency of Use of Money," loc. cit., subsequently revised.

3 From payments for final products and total adjusted deposits and currency, excluding United States government deposits, for June 30, 1909–1923, and annual averages for 1924–1947 computed from midyear and year-end data, *Banking and Monetary Statistics*, pp. 34–35 and *Federal Reserve Bulletin*. These ratios run lower than those in the preceding column because adjusted total deposits and currency includes, while the quantity of money held by business and individuals excludes, deposits of State and local governments and of foreigners. The lack of uniformity in the difference between the two series is due in part to the same reason, and in part to the difference between (a) figures at mid-year or average of mid-year and year-end figures and (b) those data adjusted on the basis of available weekly and monthly information.

4 From sources described in preceding footnotes.

5 As of June 30 for 1909–1923; annual averages for 1924–1947 computed from mid-year and year-end data, *Banking and Monetary Statistics*, pp. 34–35, and *Federal Reserve Bulletin*.

expenditures for final products to the quantity of money from 1909 to 1939, and a trend line at 1 1/3 per cent decrease per year. The dotted portion of the line showing the ratio of national income to quantity of money covers a period when the banking data are deficient (see footnote 6 to Table II). These data indicate that the downward trend in monetary velocity is not of recent origin, since it apparently extends back a century and a half.

### THE TREND IN VELOCITY AND GROWTH IN TIME DEPOSITS

The long-term downward trend in the circuit velocity of money represents a tendency of business and individuals to hold, as time goes by, a larger quantity of money relative to their expenditures for the output of the economy. It would be natural to expect that a substantial portion of these relatively larger cash balances would be held in the form of time deposits; that is, in that form of money which emphasizes the store of value function. Payments for final products of the economy are made chiefly by use of demand deposits or currency, and it is desirable to compare ratios of such payments to the sum of demand deposits and currency, with the ratios to the total quantity of money, or to the total amount held by business and individuals.

FIGURE II

CIRCUIT VELOCITY OF MONEY INCLUDING AND EXCLUDING TIME DEPOSITS
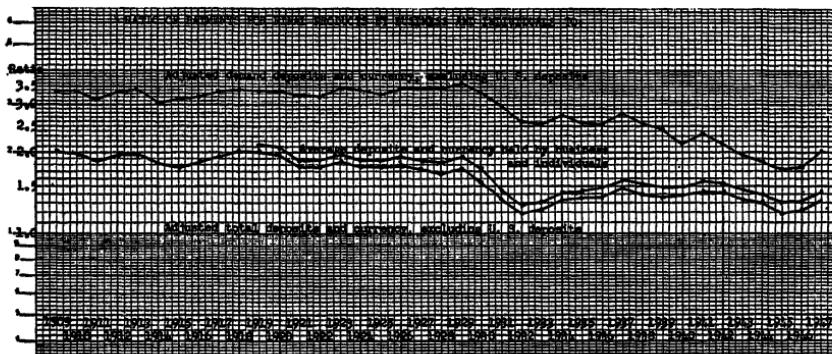


Estimates of demand deposits adjusted and currency outside banks, prepared by the Board of Governors of the Federal Reserve System, go back to 1892. However, annual estimates of expenditures for final products are available only as far back as 1909. In Figure II annual ratios of payments for final products to adjusted demand deposits plus currency outside banks are compared, for the period

since 1909, with the ratios to the total quantity of money. Various differences between the rates of use of money calculated on the two bases are evident. Figure III, which shows the relative importance of

FIGURE III

RELATIVE IMPORTANCE OF TIME DEPOSITS IN THE TOTAL QUANTITY OF MONEY
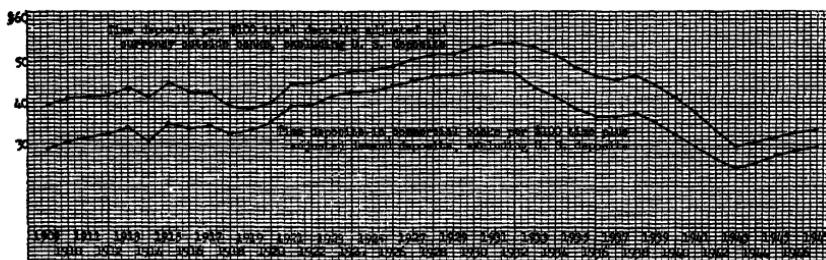


time deposits in the quantity of money, throws considerable light on these divergences. The data used in these charts are given in Table III.

In the nineteen-twenties the downward trend in the rate of use of money, as calculated from the total quantity of money, does not appear in the ratio of payments for final products to demand deposits and currency. In fact, from 1921 to 1929 there is an upward trend in the ratio of payments for final products to adjusted demand deposits plus currency outside banks. This difference in trend is associated with a rapid growth in time deposits. From 1919 to 1931 time deposits increased at a much faster rate than demand deposits and currency, rising from 39 per cent of the total quantity of money other than United States government deposits, in the former year, to 55 per cent in the latter.

This large growth in time deposits relative to the growth in demand deposits and currency was the result of provisions of the Federal Reserve Act which affected the classification of deposits by member banks. Prior to the establishment of the Federal Reserve System, no differentiation was made between demand and time deposits with respect to reserve requirements for national banks, and the same was true in the majority of States with respect to State banks.[6] The Federal Reserve Act prescribed a much lower percentage for time deposits than for demand deposits. Subsequent to 1917, when bank reserves were concentrated in the Federal Reserve System,

6. For State legislation see Samuel Wellden's *Digest of State Banking Statutes*, prepared for the National Monetary Commission, Senate Document No. 363, 61st Congress, 2d Session (1910).

member banks — in order to expand their earning assets as much as possible with the reserves they possessed — found it advantageous to encourage their depositors to shift their customary minimum balances in checking or other demand accounts into accounts labeled "time."

The tendency to reclassify demand deposits into time deposits was reversed in 1933, when withdrawal of time deposits without advance notice became more difficult. This change in the law, together with the growth of service changes on checking accounts, stimulated a reclassification of deposits in the opposite direction from that which occurred during the nineteen-twenties. During most of the period since 1933 time deposits have declined relative to total deposits or relative to total deposits plus currency. As a result, the downward trend in monetary velocity shows up not only in ratios of payments for final products to total money, but also in ratios to demand deposits and currency. In fact, the calculated downward trend in monetary velocity has been more severe when computed on the basis of demand deposits plus currency, than when computed on the basis of total deposits plus currency.

Prior to establishment of the Federal Reserve banks, as indicated by the data from 1909 to 1915, the situation was intermediate between that of the nineteen-twenties and that subsequent to the early nineteen-thirties. From 1909 to 1915 time deposits were growing more rapidly than demand deposits and currency; and the ratio of payments for final products to demand deposits and currency appears to have no significant trend.[7] For this period the data suggest that the increased holding of money relative to expenditures was largely in time deposits.

From 1915 to 1919 the situation is more complex. For these years the downward trend in the ratio of payments to money was interrupted, not only in the case of the ratio to demand deposits and currency, but also in that of payments to the total quantity of money. This may have been due in part to wartime conditions which affected the usage of money. However, another factor responsible for this

7. If a trend exists, it is obscured by the irregularity of the ratios. Further, it is difficult to judge whether the irregularity of the ratios of payments to demand deposits and currency is primarily cyclical, or is due chiefly to inaccurate estimates of the respective amounts of demand and time deposits. The estimated amounts of time and demand deposits for 1909 are based on a special report collected by the National Monetary Commission, and those for 1915 chiefly on reports submitted by member banks. For the intervening years the amounts of demand and time deposits are based on a somewhat incomplete classification of deposits. In several of these years certificates of deposits in State banks are not separated into time and demand certificates; and in certain years savings deposits in national banks are not adequately segregated from individual deposits subject to check.

interruption to the downward trend in the rate of use of money, was the introduction of the Federal Reserve clearing system. This nationwide clearing system was a notable improvement in the method of making payments, and, by increasing the efficiency of money, lifted the level of ratios of payments to the quantity of money, however measured.[8]

The foregoing data make it clear that calculations of monetary velocity made on the basis of demand deposits plus currency have been greatly affected by factors which influenced the classification of deposits without altering their use. For this reason such measures of monetary velocity are less significant for general economic analysis than measures of monetary velocity based on the total quantity of money including time deposits.

Another aspect of the variability of time and demand deposits, relative to each other, is the fact that additional supplies of money are injected into the economy almost exclusively in the form of demand deposits. Transformation into time deposits of a portion of an increment to the money supply typically occurs somewhat later than the date of the increment to demand deposits. It is reasonable to assume that in part this transformation is a formal recognition by depositors of their increased cash reserves — that is, that the deposits had in effect become time deposits *prior* to their transfer to time accounts. Because of this lag, measures of monetary velocity based on demand deposits and currency — even if they could be corrected for changes in classification of deposits induced by changes in banking codes and in service charges — would not be as useful a tool for use in analysis of the relation of banking to business fluctuations, as are measures based on total deposits and currency.

### SEASONALITY OF PRODUCTION AND THE RATE OF USE OF MONEY

One of the important influences upon habits of use of money is the periodicity of payments associated with seasonal production. However, a high seasonality in production and in the receipt of personal income will result in a comparatively low or in a comparatively high rate of use of money, depending upon the degree of mercantile credit extended to income recipients, and upon correlative interrelations of mercantile and bank credit. Assume, for example, that mercantile credit to individuals is very small, and that merchants

8. The data used in Figure 1 suggest that another interruption to the downward trend in monetary velocity may have occurred in the eighteen-forties, but the monetary data are inadequate and this conclusion is not warranted. See footnote 6 to Table II.

and dealers do not borrow extensively from the banks. Bank assets therefore consist chiefly of government obligations or other long-term assets, and may be assumed to have little seasonality. Assume also that consumer expenditures are made evenly throughout the year, and consider the extreme case of income receipts concentrated on one day in the year. On the morning of that day individuals have negligible cash balances and virtually the total quantity of money is in the hands of traders and other business concerns. At the close of the day the money of the nation is held by individuals. During the year money steadily moves back into the hands of traders and business concerns. The annual ratio of payments for final products to the quantity of money is 1; and the reciprocal of this ratio, which measures the time taken by the average unit of money to make its round from consumers to business concerns and back again, is also 1 (i.e., one year). This is a very slow rate of use of money, compared with the same kind of banking system, but with frequent periodic income payments. If incomes, for example, are received regularly twice a month, and spent evenly during the period from one payday to another (with the money supply on the morning of payday entirely in the coffers of traders and business concerns, and in the evening of that day entirely in the pockets of income-recipients), the annual ratio of payments for final products to the quantity of money is 24; and its reciprocal, measuring the average circuit time of money, is 1/24 of a year.

However, the calculated velocity of money, under a system of income payments once a year, will not be so slow if retail sales are made on credit — with accounts cleared up once a year when incomes are received — and with the assets of the banking system consisting of "commercial" loans. For illustration, let us take two extremes — the first in which the custom of credit sales, with annual clearance of indebtedness, extends throughout the economy; the second in which this custom is confined to sales of final products.

In the first case no one in the economy will have any need for money (bank deposits or currency) except at one period of the year. Let us suppose, therefore, that on the day preceding that on which all incomes are received, employers and dealers in farm produce make short-term bank loans; and that the various stages in the clearance of all accounts are made on successive days. The process is as follows: on one day, bank loans are made and the money supply created; on the next day, income payments are made; on the third, individuals pay their debts to retailers; on the fourth, retailers settle with jobbers and other intermediate dealers; on the fifth, all intermediate dealers

pay their debts to business concerns; and on the sixth, all business concerns repay their bank loans. The money supply of the nation — whether measured by the deposit and currency liabilities of the banking system, or by the cash balances held by individuals — is zero on 360 days in the year, and equals the aggregate expenditures for final products on five days. The annual ratio of expenditures to the average money supply is 72; and the calculated average circuit time of money is 1/72 of a year, or six times a month.[9]

In the second case of sales of final products on credit with annual clearance of indebtedness, but with this custom confined to such sales, we assume that mercantile borrowings constitute the assets of the banking system. Merchants clear up their bank loans once a year, when they are paid by their customers; but in order to restock their shelves as they make sales during the year, they gradually increase their bank loans from zero just after settlement day to the amount of their year's sales just before settlement day. The deposits and currency created as these loans are built up accrue throughout the year in the coffers of business concerns and are paid over to individuals on income payments day. The average outstanding quantity of money is one-half the annual payments for final products, the ratio of such payments to the money supply is two, and the average circuit time of money is one-half of a year.[1]

## FACTORS IN THE DECLINING TREND IN RATE OF USE OF MONEY

The factual data, as has been indicated above, show that the declining trend in rate of use of money, as indicated by a ratio of the value of the nation's output to the quantity of money, is of long standing. Such a permanent economic phenomenon must have an explanation in terms of the conditions under which money is used in purchasing the output of the economy. It is desirable, therefore, to take notice of various social and economic factors which are presumably responsible for the declining velocity of money. Several factors, including the effects of changes in the seasonality of production, are mentioned here. These factors are not all independent of each other. In fact, some of them are definitely concomitant, and their effect upon the rate of use of money is a joint effect.

9. If the settlement week is in October, and the date for which information on bank assets and liabilities is reported is June 30, the circuit velocity of money, calculated from the available statistical data, becomes infinite.

1. If settlement day is October 31, and bank asset and liability data are collected for June 30 only, the ratio of payments for final products to the money supply, calculated from the statistical data, becomes 1.5 and the circuit velocity of money two-thirds of a year.

One of the factors in the apparent decline in rate of use of money, particularly during the nineteenth century, is the increase in the proportion of the nation's output passing through the market economy. Although estimates of national income and of the value of the nation's output throughout the nineteenth century are not sufficiently detailed or reliable to show how large a proportion of the goods and services represented by that value was sold in the markets, and how large a proportion was produced by farm and village families and other persons for their own use or for barter, there is no doubt that the latter proportion was substantially larger in the earlier than in the later part of the period covered by Figure I. Information regarding exchange of products through barter rather than through the use of money is exceedingly scanty. However, the literature of the nineteenth century indicates that barter was more extensively used in the early part of the century than in more recent times.

Farm products are a smaller proportion of the nation's output than formerly, and are now sold more evenly throughout the year. These changes have reduced the proportion of the income of the people received in relatively large sums at one period in the year, when crops are sold. These developments have affected not only the income of farm families relative to that of wage and salary recipients, but also the income of traders handling farm products. Taken alone, such a change would result in a more rapid use of money; but taken together with changes in the extent and terms of mercantile credit, the result has probably been a reduced annual rate of use of money for purchase of final products.

Contrary to a prevalent impression, there appears to have been, over the period of the past century and a half, a substantial reduction in the length of time for which credit is typically used in the purchase of final products. References to the custom of paying all bills once a year are found in literature relating to the early part of the nineteenth century.[2] Only for durable goods is a year's credit now customary.

Also, contrary to a prevalent impression, the proportion of consumers' goods and services sold for cash relative to the portion sold on credit (apart from the length of term of credit) has probably increased substantially in recent decades relative to former times. The economic and social literature of the last century contains references to the credit extended by food stores to working men as well as to other classes of the population; the predominance of "cash only" food stores is a modern development. Since expenditure for

2. I have not collected a file of such references, nor of those mentioned in the next paragraph.

food is a large percentage of total expenditure for final products, the increased proportion of food sales for cash, and concomitant changes in credit terms throughout the food handling industries, and in the use of bank loans by the food industry, from grocery stores on back to farmers, may have had an appreciable effect upon the trend in rate of use of money.

It appears probable also that the percentage of other consumer expenditures made for cash has increased. Transportation expenses of individuals, for example, formerly consisted largely of horse feed, also commonly purchased on account; in recent years they consist chiefly of cash purchases of gasoline and other operating automobile expenses, and of cash fares for public utility transportation. Further, current expenditures for transportation have probably become an increasing proportion of total expenditures for final products. Use of credit for purchase of automobiles has, of course, been large in amount, but whether it is greater or less, relative to total consumer expenditures, than credit purchases of horses and wagons in the nineteenth century is unknown.

A tendency may have occurred, particularly during the last half-century, toward longer wage-periods. If income is received periodically but is spent evenly throughout the period between paydays, the longer the interval between paydays the larger the average holdings of cash relative to annual expenditures. Definite information is not available, but it appears likely that a substantially smaller proportion of urban and village wage-earners are paid by the day than in former times, and also that a larger proportion of workers in industrial establishments are paid once or twice a month, or bi-weekly, relative to the number paid weekly.

A larger proportion of the population than formerly work for wages or salaries rather than in enterprises under their own control. In view of the insecurity of employment, and therefore of wage income, this has probably tended to raise the average amount of cash held as a reserve for bad times, periods of unemployment, or emergencies.

More intermediate payments may be involved in the process of production than in former times because of more complex or "roundabout" methods, or greater specialization of business organizations. A larger number of intermediate payments would be expected to require a larger number of cash reservoirs, and therefore result in the holding by business enterprises of a larger amount of cash relative to the final value of output.

It should be noted also that vertical integration of industry does not necessarily reduce the number of payments involved in con-

nection with the productive process for an item of final output. Accounts of formerly separate establishments may be handled after integration in such a manner that payments by check are made by one establishment to another, even though the establishments are part of the same business firm. Integration of merchandise concerns, such as the expansion of chain grocery stores, may also increase the amount of cash balances held by trading concerns relative to the value of sales to consumers. This may result from the fact that each local outlet maintains an account with a local bank and transmits the proceeds to central organizations, thus involving two sets of cash balances for a given volume of sales. Such an increase in the amount of cash holdings relative to sales may or may not be offset by more direct purchasing by the headquarters of the chain.

Higher rates of estate taxation, together with the difficulty of disposing of large amounts of property at a given time, have tended in recent decades to an accumulation of cash balances by elderly people. This tendency may have been strengthened by attempts to evade taxation through placing assets in a form most easily passed on to heirs without reporting.

During the past century a great change has occurred in the nature of the investing process, accompanying the growth of large corporations and the transformation of the "middle classes" from independent or professional businessmen or entrepreneurs to salaried employees. Decades ago, the "middle classes" tended to invest the cash they had available for this purpose in their own business enterprises, or to participate directly with neighbors or business acquaintances in the development of other growing enterprises or new ventures. Investments by the middle class today must be made largely in corporate stocks or bonds, or through financial intermediaries such as insurance companies; and there is probably a tendency to accumulate cash balances for investment purposes. These developments may have resulted in relatively larger holdings of cash awaiting investment than was formerly the case.

Some contemporary economists believe that investors now hold cash balances awaiting investment in relatively larger amounts than formerly because of uncertainty regarding interest rates or because of a lowering of the level of interest rates. The latter does not appear to be an important factor in the secular trend toward larger cash holdings relative to the value of the output of the economy, since the trend in monetary velocity has persisted through periods of advancing interest rates, as in the first decade of this century. Uncertainty regarding monetary policy, and consequent uncertainty regarding

business prospects and future interest rates may at times have had a considerable effect on the holding of relatively large cash balances. Such an effect, however, is associated with periods of depression and deviations in monetary velocity below trend, as in the nineteen-thirties, rather than with the long-term downward trend in velocity.

In addition to the foregoing, there is probably a general tendency for people to hold larger cash balances, as time goes by and their average income increases, relative to their expenditures for goods and services. This is in line with the tendency to hold larger stocks of other types of goods and services, such as the number of shoes or suits of clothing relative to the number in use at a particular moment.

The relative importance of the foregoing factors cannot be adequately appraised. It is evident, however, that the net effect of social and economic influences such as those mentioned has caused an upward drift over time in the average holdings of cash balances by individuals and business enterprises relative to the value of the final output of the economy. This drift, or secular trend, judged by available but not fully adequate evidence, appears to have been maintained over a longer period of time than any of the other secular trends recorded in economic data, excepting those of the growth of population and of production.

Whether such factors as the foregoing will continue to operate in the future in the degree to which they have operated during the last century and a half may be questionable. Some of the factors enumerated have probably lost most of their force, others are probably as strong today as at any previous time. Pending factual evidence that the declining secular trend in the general rate of use of money is weakening, the assumption should be made — as part of the basis of current monetary policy designed to provide full use of resources or full employment — that the trend will continue. The increase of about one and one-third per cent per year in the quantity of money which is needed because of this trend is small enough — in comparison with the growth in the quantity of money needed on account of the trends in population and productivity — so that monetary policy will not be seriously in error if detection of a change in the trend in rate of use of money lags behind its occurrence.

## INTERRELATIONS BETWEEN CHANGES IN THE QUANTITY AND VELOCITY OF MONEY

The other controversial issue respecting monetary velocity mentioned above is the comparative timing and the interrelations of deviations in the quantity and in the rate of use of money. A widely

prevalent assumption is that business recessions start with a decline in aggregate expenditures which arises from causes other than monetary policy, and which therefore involves a decline in the rate of use or velocity of money. Relevant statistical data bearing on this question have been examined elsewhere and a brief resumé of the conclusions which have been reached is given here.[3]

Factual information for the period since 1919 does not support the assumption that variations in monetary velocity are an initial factor in business depression. The data do, however, indicate that in some cases a declining velocity of money has accompanied and in other cases has followed downward deviations from trend in the quantity of money. In fact, after a business recession has run for a time and the quantity of money has been reduced, there is almost uniformly a slowing down in velocity, relative to trend, which is reversed only when the shrinkage in the money supply is known, or believed, to have been stopped. There is no evidence that disturbances to economic equilibrium originate in an erratic rate of use of money, but there is much evidence that such disturbances result in, and are in turn intensified by, variations from trend in the rate of use of money.

<div align="right">CLARK WARBURTON</div>

McLEAN, VIRGINIA

3. "Quantity and Frequency of Use of Money," loc. cit.; "Bank Reserves and Business Fluctuations," *Journal of the American Statistical Association*, Vol. 43 (December, 1948), pp. 547-58.

# THE INTERNAL ORGANIZATION OF THE FIRM AND PRICE FORMATION: AN ILLUSTRATIVE CASE

## I. INTRODUCTION

If any single conclusion can be drawn from the recent marginalist controversy as it relates to the price policies of business firms,[1] it is that an understanding of price formation must be derived in large measure from the study of actual market structures. Although it is significant that, in the controversy referred to, anti-marginalists as well as marginalists employed a common theoretic frame of reference, it seems no less significant that much of the argument on both sides turned on the appeal to market facts. Hall and Hitch, for example, while believing that their study "casts doubt on the general applicability of the conventional analysis of price and output policy in terms of marginal cost and marginal revenue," cite, as a reason for the adoption of a "full cost" price policy, the fact that "changes in price are frequently very costly, a nuisance to salesmen, and are disliked by merchants and consumers."[2] This is an explanation of price policy in terms of market structure. Machlup, on the other hand, concluding that "the marginal theory of business conduct of the firm has not been shaken, discredited or disproved by the empirical tests" of anti-marginalists, nonetheless points out that government regulation, accounting convention and the nature of cartel administration — among many other causes — are responsible for the adoption of the average cost pricing policy.[3] This, again, is an explanation of price policy in terms of market structure.

1. Recent discussion of marginal theory and price policy includes: R. L. Hall and C. J. Hitch, "Price Theory and Business Behavior," *Oxford Economic Papers*, No. 2 (May 1939), pp. 12–45; Fritz Machlup, "Marginal Analysis and Empirical Research," *American Economic Review*, XXXVI (September 1946), pp. 519–554; and Henry M. Oliver, Jr., "Marginal Theory and Business Behavior, "*American Economic Review*, XXXVII (June 1947), pp. 375–383, and "Average Cost and Long-Run Elasticity of Demand," *Journal of Political Economy*, LV (June 1947), pp. 212–221. The match which rekindled discussion of empirical tests of marginalism was applied by Richard A. Lester in an article on employment policies: *American Economic Review*, XXXVI (March 1946), pp. 63–82.

2. Op. cit., pp. 12, 22.

3. Op. cit., pp. 538–543, 553.

Considerations such as these may have led Oliver to suggest that "very likely both illogic and incomplete adherence to marginal reasoning form parts of an accurate explanation."[4] Business men adhere to an "incomplete marginalism." The terms of their adherence and circumstances conditioning the actual formation of price are to be sought through investigation of particular market structures.

Development of the theory of monopolistic competition in the decade before the war revealed possibilities for a wide range of price policies open to business firms.[5] It became evident, also, that price policies were determined not only by the numbers of sellers in a market and by differentiation of the product but by a great many other factors in the economic environment, some of which were internal, and others external, to the firm. Economists became concerned about how the processes of price formation within the firm were affected by such factors as its size, its organizational structure, and its ownership. Economists also looked to external features of market organization such as the nature of the product, industry cost structure, age of the industry, technological background, ease of entry and disinvestment, and the nature of distributive channels. These and similar considerations became the principal objects of study in the search for determinants of price.[6]

Case studies in price policy have tended to concentrate in agriculture and in manufacturing and extractive industry.[7] In these

4. "Average Cost and Long-Run Elasticity of Demand," loc. cit., p. 218.

5. E. H. Chamberlin, *Theory of Monopolistic Competition* (1st ed.; Cambridge, 1933). With reference to the marginalist controversy, it may be noted that Professor Chamberlin, although presenting the marginal curves as an alternative "graphic device" (p. 14) for describing equilibrium, makes little specific use of them in his analysis, and argues (5th ed., p. 192) against exaggerating their importance. Also, what has come to be called the "full cost" principle, which is generally regarded as opposed to "marginalism," and is described by Hall and Hitch (loc. cit., p. 30) as a departure from "what a reader of Professor Chamberlin's book would infer," is, in fact, a part of his analysis (p. 105).

6. E. S. Mason, "Price and Production Policies of Large-Scale Enterprise," *American Economic Review*, XXIX Supplement (March 1939), pp. 61–74; Joe S. Bain, "Market Classifications in Modern Price Theory," this *Journal*, LVI (August 1942), pp. 560–574.

7. E.g., Joe S. Bain, *The Economics of the Pacific Coast Petroleum Industry* (Berkeley, 1944–1945); John D. Black, *The Dairy Industry and the AAA* (Washington, 1935); John M. Cassels, *A Study of Fluid Milk Prices* (Cambridge, 1937); C. R. Daugherty, M. G. de Chazeau and S. S. Stratton, *The Economics of the Iron and Steel Industry* (New York, 1936); John A. Guthrie, *The Newsprint Paper Industry* (Cambridge, 1941); A. C. Hoffman, *Large-Scale Organization in the Food Industries* (U. S. Temporary National Economic Committee Monograph No. 35; Washington, 1940); John P. Miller, "The Pricing of Bituminous Coal; Some International Comparisons," *Public Policy*, ed. by C. J. Friedrich and Edward S. Mason (Cambridge, 1940), pp. 144–175; Lloyd G. Reynolds, "The Canadian

fields certain aspects of external market structure and internal organization of business units have led to distinctive patterns of price policy such as the basing-point system, price leadership, zone pricing, base rating, and price stabilization. These studies have investigated the manner in which certain price policies have arisen from the facts of the economic environment. Many of the studies have also proceeded to assess the effects of these price policies on public welfare.

Few case studies of this sort have been attempted in the field of retail trade. Several investigators have presented general studies of retail price formation by the use of techniques derived from the theory of monopolistic competition,[8] but at this juncture what most needs to be undertaken is the investigation of competition in particular retail markets through a series of case studies. The present paper is an attempt to develop such a study of price formation in the department-store market by concentrating attention on factors of internal firm organization. These factors are: the large size of the store unit, administrative organization, budgetary procedures, interunit controls and owner-management relationships. In a final section certain price policies resulting from these conditions are reviewed.

## II. LARGE SIZE OF THE STORE UNIT

Price policies in department stores are influenced, in the first place, by the large size of the store unit. Because it combines under one roof and under one management the facilities of many shops handling various lines of merchandise, it is necessarily very large. The United States Bureau of the Census defines a department store as a retail distributor of general merchandise with annual sales in excess of $100,000. That this required minimum sales volume is well above that of the average retail store can be seen from Table I. It will be noted that over half of all retail stores had less than $10,000 sales per unit in 1939. Department stores are frequently much larger

Baking Industry," this *Journal*, LII (August 1938), pp. 659–678; Homer B. Vanderblue, "Pricing Policies in the Automobile Industry," *Harvard Business Review*, XVII (Summer 1939), pp. 385–401, also XVII (Autumn 1939), pp. 64–81; Donald H. Wallace, *Market Control in the Aluminum Industry* (Cambridge, 1937).

8. Henry Smith, *Retail Distribution* (London, 1937); Ewald T. Grether, *Price Control under Fair Trade Legislation* (New York, 1939), pp. 225–255; E. R. Hawkins, "Marketing and the Theory of Monopolistic Competition," *Journal of Marketing*, IV (April 1940), pp. 382–389; John F. Due, "A Theory of Retail Price Determination," *Southern Economic Journal*, VII (January 1941), pp. 380–397; W. Arthur Lewis, "Competition in Retail Trade," *Economica*, XII (November 1945), pp. 202–234.

than the minimum size of $100,000 in sales volume; in 1939 average sales per store were $975,000.

TABLE I

Retail Stores by Size Groups, 1939

| Annual Sales | Number of Stores | Per Cent of Total | Sales (add 000) | Per Cent of Total |
|---|---|---|---|---|
| $300,000 and over....... | 12,630 | .7 | $9,855,631 | 23.4 |
| $100,000–$299,999...... | 50,097 | 2.8 | 7,955,285 | 18.9 |
| $50,000–$99,999........ | 93,318 | 5.3 | 6,394,703 | 15.2 |
| $30,000–$49,999........ | 133,221 | 7.5 | 5,077,007 | 12.1 |
| $10,000–$29,999........ | 522,117 | 29.5 | 8,938,632 | 21.3 |
| Less than $10,000....... | 958,972 | 54.2 | 2,820,532 | 9.1 |
| | 1,770,355 | 100.0 | $42,041,790 | 100.0 |

Source: U. S. Bureau of the Census, *Retail Trade: 1939,* Pt' 1, p. 48.

The large absolute size of the department store (as contrasted with its large relative size in particular retail markets) necessitates extended delegation of authority to deal with pricing. Although departmental grouping of similar types of merchandise permits the application of common pricing principles to broad categories of goods, there is nonetheless a multitude of individual decisions on price to be taken in a business which consists basically in the buying and selling of goods. R. H. Macy and Company, for example, carries over "400,000 separate and distinct items, not counting colors and sizes."[9] Unfortunately for simplicity of operation in a department store, the separate items are not related to one another in a formal structure of prices as in the steel industry, where management can alter the entire price structure by changing a relatively small number of base prices. Thus, the very great number of prices and the absence of a formal price structure force major executives to delegate much pricing authority to lower levels. The immediate effect of this diffusion of pricing jurisdiction is to substitute for the direct response of the individual merchant the routine of a large organization.[1]

Large size has also directly influenced the selling methods adopted by the department store. Certain practices, such as the

9. Letter of Mr. Q. Forrest Walker, Economist for R. H. Macy and Company, to the writer.
1. "Much of the action of individuals as participating in organization is habitual, repetitive, and may be merely responsive by *organization design* — a result, for example, of specialization intended to enhance this non-logical process." Chester I. Barnard, *The Functions of the Executive* (Cambridge, 1938), p. 186.

branding of the store's own line of merchandise, are unavailable to competitors of smaller size. In this regard, the unit of real significance is the total selling organization which, in a retail chain, may include many separate stores. As another example of selling methods open only to the large store, newspapers of wide circulation in the metropolitan market area are used with greater efficiency by the department store than by smaller competitors whose geographic markets are narrower. Through its large sales volume the department store, in effect, creates special conditions of sale which act as major determinants of resale prices.

### III. ADMINISTRATIVE ORGANIZATION

A second aspect of internal structure significant for price formation is the elaborate administrative organization common to all department stores. This is partly a result of the aggregate size of the store, but it is also a function of the diversity in kinds of goods carried. The organization of the department store is built around the operations of merchandising, publicity, store management and control. These functions are ordinarily performed by separate divisions, and it is the first, or merchandising, division which has the responsibility for buying and selling. It is composed of the several merchandise selling departments, — from which the *department* store derives its name — each in charge of a "buyer"[2] who, in addition to buying and selling, also trains and directs salespersons, controls expenses, initiates advertising, and — not least important — prices merchandise. Historically, the buyer is the inheritor of the founder's tradition of all-around merchant. In terms of organizational theory he has a complete circle of responsibility for a process which has been delegated to him.

Some understanding of the pricing process in the department store can be gained by examining the position occupied by the buyer within the total administrative structure. Immediately, the buyer reports to a divisional merchandise manager who co-ordinates the work of several buyers handling related lines of merchandise. Thus, there are divisional managers for home furnishings, for women's ready-to-wear clothing, for men's and boys' wear, and for the basement store. The divisional merchandise manager supervises purchasing, the planning of sales, the setting of gross profit margins

2. The term "buyer" is misleading; actually he is more properly called a department manager, a practice which some chain organizations, who rely more heavily on central buying, have initiated. National Retail Dry Goods Association, *The Buyer's Manual* (2d ed.; New York, 1931), pp. 18–20.

(mark-up), the reduction in prices (mark-down), and the rate at which stocks are cleared (turnover). He brings to the pricing operation a statistical view which may dampen the ardor of an enthusiastic buyer, or prod a timid one to more aggressive action. The divisional manager system introduces a split in the merchant's function: the buyer, who is immediately responsible for pricing, supplies the creative enthusiasm for assembling attractive merchandise and directing its sale to the public; the divisional manager, who is further removed from selling, takes a reasoned view based on statistics of performance. Price determination, under these conditions, is likely to represent a compromise of views between the buyer, who is inclined to see the marginal possibilities of pricing particular goods, and the divisional manager, who is more likely to look on departmental operations in the light of statistical averages.

Co-ordinate with the merchandise division are the publicity division, the store management division, and the control division — staff groups which perform special services indirectly conditioning the determination of prices. The publicity division is the guardian of the store's competitive position. If it is store policy to undersell competitors (on the limited range of items which can be directly compared by consumers), the publicity division will insist on low prices not only for advertised items but generally within the store. To this end the comparative shopping bureau is frequently placed in the publicity division which is thus put in the position of being able to exert downward pressure on certain prices. The publicity division, by its activities in advertising, window display, internal store display, etc., controls many of the conditions of non-price competition which in turn set the stage for actual pricing. In its staff relationship with the merchandise division, the publicity organization works with buyers in the preparation of advertising copy. The initiative of the publicity division may even extend backwards into the merchandising organization, not only to determine what to advertise, but to decide what is to be bought for advertising and at what price it is to be sold. In similar manner the store management division conditions the terms of sale and sets the limits within which the buyer — the pricing authority — acts. The store management group has taken over from the buyer much of the task of training and direction of sales personnel, of setting sales quotas, and of transferring employees among departments. This division determines in large measure the amount of retail selling service which is offered with each transaction. By controlling selling costs, the store management division influences the limits within which the buyer must set gross margins. The control

division, likewise, in discharging its responsibility of maintaining the store's net profit position, exerts considerable indirect influence over the pricing process. The controller can urge lower expense rates (which would reduce advertising outlay or change the amount of selling and other services offered by the store), or he can insist on a conservative inventory policy (which would bring pressure on the merchandise division to take early mark-downs). In short, the quadripartite administration of the department store leads to compromises on price and other matters which are quasi-political in nature.[3]

## IV. BUDGETARY PROCEDURES

A third aspect of organizational structure which conditions the setting of prices is the elaborate budgetary control system employed by the department store. Many of the decisions taken in the formulation and application of budgets, especially those affecting the buying and selling of merchandise,[4] fundamentally determine prices. We are here interested in two types of controls: (1) those which deal chiefly with dollar aggregates; and (2) those which relate to physical units of goods.

The merchandise budget deals with dollar aggregates. It plans in advance of each season, and controls in the course of the season, the principal components of the buying and selling operation: sales, mark-downs, stocks, purchases, turnover and mark-up. These magnitudes are stated for each department in total dollar terms, except in the case of turnover which is stated as a ratio, and mark-up which is a percentage of original retail price. The planning period for the department store is a six-month season beginning in February or August. The fact that the basic constituents of price are reconsidered at six-month intervals introduces a certain amount of inflexibility into the price-making process. Interim adjustments particularly in planned sales, are frequently made, but thorough-going revision of estimates, especially mark-up, is rare.

The merchandise budget usually originates with the buyer who estimates total departmental sales month by month for the coming

3. For a parallel discussion of effects on policy of conflicting points of view within manufacturing organizations, see Conference on Price Research, *Cost Behavior and Price Policy* (New York, 1941), pp. 43–44. The comment of one reader at this point was: "When internal conflicts of interest are 'compromised' rather than solved, the final decision can only through weird coincidence be in accordance with conventional marginalism."

4. Expense budgeting and control also affect price indirectly through influencing the level and variability of costs. The nature of department-store costs and their effect on price is a subject reserved for another occasion.

season on the basis of past performance and market trends. In estimating sales there is necessarily a large amount of guesswork, especially in departments affected by fashion. The making of a sales estimate requires, logically, at least, two simultaneous sub-estimates: (1) the amount of each individual commodity which can be sold; and (2) the price at which it can be sold. It is unlikely that the buyer's thinking runs explicitly in terms of how many physical units of each type will be sold, unless perhaps he happens to be dealing with a relatively small number of high valued items. It is also unlikely that either the buyer or the divisional merchandise manager makes explicit estimates at this time of the prices at which individual goods can be sold. The making of sales estimates as dollar aggregates not specifically built on estimates of sales for individual items submerges the concept of price. In preparation of sales estimates in the department store there is little evidence of the explicit treatment of price-quantity relationships in the schedule sense of economic theory.

Estimated departmental mark-downs are also stated in aggregate terms and these are based largely on historical experience. In planning mark-downs, department-store management, of course, is aware of the positive correlation of price and mark-downs (reflecting a negatively inclined demand curve). With a very low price the mark-down rate would approach zero. Thus, implicitly, by the adoption of a particular mark-down rate, management postulates a certain (although unspecified) level of price, and a certain elasticity of demand. It seems probable, however, that, in formulating a merchandise plan, management treats mark-downs as a resultant of other price-making decisions such as the estimate of sales and the rate of mark-up, rather than as an independently determined factor in price.[5]

With sales and mark-downs planned, the next move is to set the probable amount of stock on hand at the opening of the season and to estimate the amount required at the beginning of each of the succeeding five months in order to achieve the estimated volume of sales. Determination of these successive stock positions has two consequences: on the one hand it makes possible on the basis of the three parameters — sales, mark-downs, and stock — the calculation of purchases required to maintain given stock positions; and, on the other, it indicates a certain rate of turnover of the asset merchandise. This second result is closely related to the quantity of retail services offered the consumer because, in setting turnover, the buyer decides

5. Further evidence of this view is offered by the *dependent* nature of adjustments in mark-down as the season proceeds. See p. 102.

in effect that only those goods will be offered in his assortments which, on the average, are capable of maintaining a certain rate of stock turn. Setting a turnover rate, therefore, amounts to determining one dimension of price, i.e. the amount of service product (in this case consisting of "assortment utilities") to be made available at prevailing prices.

The mark-up on any one item is the difference between the initial retail price at which the good is offered and the wholesale cost, expressed as a percentage of initial retail price. A weighted average of successive mark-ups obtained on the flow of goods entering a department's stock in any given season is, therefore, a summary expression, in percentage terms, of the initial price of retailing services offered by the department. A plan for such an average mark-up might be looked on as a focus for price-making forces. At best, however, it is a blurred focus. The reasons for this are twofold: (1) the setting of planned mark-ups is not explicitly related by management either to cost or demand factors confronting the department; (2) planned mark-ups are price determinants only to the extent that they are translated into actual prices in the course of a season. Thus, management can destroy the significance of planned mark-ups either by setting them uneconomically (without direct relation to cost and demand) or by failing to use them as guides to subsequent operations. The latter point is treated below.[6] With regard to the former, it has frequently been stated that retail stores set mark-up with reference to the average cost of doing business, including a "normal" profit.[7] There can be no doubt that department-store managements are aware that a certain percentage of mark-up is mathematically necessary if expenses, including pro-rated overhead, and profit are to be covered at the planned level of sales and mark-downs. Deriving such a figure is a purely arithmetical process. Such evidence as is available on department-store practice indicates that departmental mark-ups are not set with strict regard for coverage of average costs. Moreover, although the pricing of individual items usually takes

6. Page 102.

7. "The basic pricing policy most commonly used by merchants is the average mark-up policy. To the invoice price of goods is added a mark-up that is calculated to cover the merchant's other costs, and his desired profit, and allow for mark-downs." Hawkins, op. cit., p. 382. Cf., also, John F. Due, *Intermediate Economic Analysis* (Chicago, 1947), pp. 200–201. Due acknowledges, however, that "some variation from the process of determining [store-wide or departmental] mark-up percentage . . . results from the influence which mark-up percentages used by competitors will have on the firm's policy. . . . In some fields, more or less customary mark-up percentages are used; the firms do not recalculate mark-up at frequent intervals." Hawkins recognizes similar exceptions.

competitors' prices into account, the reflection of competitors' actions on total departmental mark-ups (at least in departments where there is a wide range of merchandise) is a consideration absent, for the most part, from management's planning. Except in rare cases where severe competition brings the issue to the fore, competitive adjustments are not generally made at this level in department-store planning. The usual method of setting planned rates of mark-up leans toward the conventional adoption of the previous year's mark-up figure or the averages compiled by the Controllers' Congress of the National Retail Dry Goods Association. Where last year's actual mark-up rate is adopted there is, of course, implicit recognition of price-quantity relationships, but this procedure can best be described as the acceptance of existing price, not price determination. Perhaps the most surprising aspect, to an economist, of the department store's budgeting is the almost casual manner in which the mark-up percentage is reached. The main emphasis in the merchandise budget relates to determination of sales and stocks rather than of price.

The merchandise plan, now completely formulated, goes for approval to the general merchandise manager, where there may be more or less adjustment of the figures in the light of the latter's judgment. Once approved by the general merchandise manager, the plan goes to the controller as the store's plan for deriving certain gross revenues in the course of the next six months' season. Simultaneously, there come to the controller's desk expense estimates of the store management division. The sales promotion budget for the publicity division, which, like the store management division, has geared its budget to the basic merchandise budget, also comes to the controller, who then attempts to bring the three magnitudes — gross revenue, operating expenses, and publicity outlay — into such relation to each other as to derive the greatest net profit. Adjustments by the controller may call for greater or less gross margin and therefore may affect price policy by raising or lowering estimated mark-up. The controller's adjustments may also affect the size of the publicity budget and thereby policies of non-price competition. Similarly, insistence by the controller on reducing the store management budget serves to decrease the level of retail services and thus modify the store's service product.

The completed merchandise plan is placed in operation with the opening of the season, and even in anticipatory buying and pricing of stock before the beginning of the new period.[8] Periodic weekly

8. For an illustration of a merchandise budget, see Table 1, p. 33, in the *Merchandise Control Manual* of the Merchandise Managers' Division, National Retail Dry Goods Association (New York, 1931).

or monthly reports inform both buyer and merchandise manager of current conditions of sales, stocks, and purchases. Planned purchases which have not yet been made constitute the buyer's "open-to-buy" — and by this magical figure, far more than by any other information which can be statistically summarized, he guides his operations in the course of the period. Mark-downs and mark-ups are also tabulated, and a rise in the former above plan or a drop in the latter below plan call for corrective action on the part of the divisional and general merchandise managers. Of necessity, mark-downs are taken whenever it is recognized that merchandise is not selling. The fact that the actual mark-down rate exceeds that in the plan will not, of course, prevent the taking of further mark-downs. On the other hand, failure to maintain planned initial mark-up will not be looked upon in the same light: the planned mark-up is a standard for performance, even though the criteria for setting the standard are somewhat ill-defined. Mark-up on any given invoice, or on some of the items on an invoice, may, to be sure, fall below plan, but over the season the buyer is under pressure to maintain the full required percentage of mark-up. For the buyer, planned mark-up imposes the requirements of an average — there may be dispersion about the average, but, short of difficulties in justifying unusually high and unusually low mark-ups to his superiors, the range of variation is one for him alone to determine.[9]

While the *general* level of prices within a department is thus determined by the merchandise plan developed in aggregate dollar terms, many aspects of individual price determination are more evident in the unit merchandise controls where the budgetary process runs in terms of physical units of particular commodities. Many goods within a department store cannot be satisfactorily subjected to the processes of unit controls; these are usually items of small unit value where the unique character or the style factor makes controls impracticable. Art goods, embroideries, fancy linens are among the items which fall into this class. On the other hand, for the vast bulk of department-store merchandise, unit controls of sales and stock are useful and are, in fact, employed.

The use of unit controls calls for the classification of goods into commodity groupings within the department; and, in each commodity

9. In industrial markets it has often been found that the pricing process starts with consideration of costs. After determining the price suggested by cost, it is appropriately revised by other considerations such as expected volume of sales, prices, competition and general business conditions. Cf. Conference on Price Research, *Cost Behavior and Price Policy*, p. 46. Pricing in the department store starts with a consideration of what can be sold rather than with the cost of the product. Cf. Smith, *Retail Distribution*, pp. 23–24.

classification, for the periodical listing of each item of stock and sales in units, through either a "perpetual inventory," or other accounting method. The important fact from the standpoint of this study is that a system of unit controls, while maintained by the buyer and divisional manager for the purpose of balancing stocks, serves to canalize sales effort into particular commodity classifications and particular price lines. In other words, price making forces express themselves only in these selected categories and prices.

Unit controls are often extended into what is termed a basic or model stock plan. This is the determination of an ideal investment in units for each price line and for each classification of merchandise. Periodically, stock clerks place reorders which bring quantities back to the level of the model stock plan. Such a semi-automatic system of keeping up stocks may produce as a by-product a reluctance to upset the apple-cart by frequent price and product changes. Where this occurs delegation of authority results in a somewhat rigid system which adjusts only at intervals to changes in market conditions.[1]

## V. INTER-UNIT CONTROLS

Although the discussion thus far has centered on the organization of the individual store unit, many department stores are owned by chains which might be expected to exert control over the formation of prices within member stores. In 1939, two-thirds of all department stores were chain units, which, however, being of smaller average size than the independents, accounted for only one-third of all department-store sales.[2] Moreover, the share of business falling to chains has constantly increased, as have the number of chain units.[3]

1. The physical process of changing marked prices on merchandise in a department store offers further resistances to price change. When goods are to be raised or lowered in price, they must ordinarily be taken off sale, trucked to the marking room, a mark-up or mark-down record made out and approved for each item, new price tags made and affixed, and the merchandise returned to sale. Under these conditions it is understandable why goods may remain on the shelves for long periods at the same prices.

2. *Retail Trade: 1939*, Pt. 1, p. 15. These data probably understate the importance of chains because the Bureau of the Census counts as independent stores "three or fewer stores under one management-ownership." Ibid., p. 845. A certain amount of sales volume properly belonging to ownership groups, which have for the most part been classified as chain stores, is also allowed to slip in with independents. U. S. Bureau of the Census, *Retail Chains* (Washington, 1937), p. 23.

3. The following data are presented in *Retail Trade: 1939*, Pt. 1, p. 15:

|  | Number of Chain Department-Store Units | Per Cent of All Department-Store Sales |
|---|---|---|
| 1929 | 1964 | 18.8 |
| 1935 | 2423 | 30.3 |
| 1939 | 2672 | 34.0 |

Department-store chains are of two types: (1) integrated chains and (2) ownership groups. Integrated chains maintain a centralized or regional control over the merchandising activities of component store units, which are operated under the common corporate name of the chain (as in the case of Sears Roebuck and Montgomery Ward). Ownership groups are composed of formerly independent stores, which have been purchased (usually many years after the founding) by the present chain. For the most part they continue to operate under original names rather than under those of the owner chains. Ownership groups permit considerable decentralization of merchandising control. Frequently stores within a group diverge as to policies, price lines, type of clientele, suppliers, etc. Unlike integrated chains, ownership groups do not rely primarily on central buying; they resemble public utility holding companies in that their control, in many instances, is chiefly financial. The relative importance of these two types of chains, as well as of independents, in 1935, is shown in Table II.

TABLE II

DEPARTMENT STORE UNITS AND SALES, BY OWNERSHIP TYPES, 1935

| Ownership Type | Store Units | Sales (add 000) | Per Cent Total Department-Store Sales |
|---|---|---|---|
| Independent............ | 1711 | $1,517,538 | 51.9 |
| Ownership Group....... | 125 | 717,407 | 24.5 |
| Integrated Chain....... | 2423 | 683,101 | 23.4 |
| Unclassified............ | 34 | 6,824 | .2 |

Source: Charles F. Phillips, *Marketing* (New York 1938), p. 394. Although the Bureau of the Census does not distinguish between the two types of chains, Phillips, by manipulating Census data for 1935, has been able to arrive at the above results. As he points out—the number of independents and chain units is overstated. The amount of the overstatement appears to be 125, the number of ownership group units, but it is not clear how the overstatement is distributed between independents and chains.

Although there is little evidence as to the distribution of pricing authority between headquarters of chains and store units, a study by the Federal Trade Commission in 1929–30 showed a surprising delegation of control on the part of chains to their store units. These results are summarized in Table III. Ownership group chains probably permit greater local autonomy than do integrated chains; but reports to the Federal Trade Commission of Ward, Sears and Penney — all integrated chains — also indicate complete freedom in meeting competition, in pricing for local trade, and in marking down merchandise which for some reason does not sell.[4] The rather meagre evidence available points toward the conclusion that price-making

4. U. S. Federal Trade Commission, *Chain Stores: Chain-Store Price Policies* (Washington, 1930), pp. 36–39.

forces have their seat in the local store unit and that price policies are not generally transmitted from central offices. On the whole we might expect this condition to result in more flexible pricing adapted to local markets.

TABLE III

DELEGATION OF CONTROL OVER PRICES AND MARK-UP
BY DEPARTMENT-STORE CHAINS TO LOCAL STORE UNITS, 1929–1930

|  | Prices | Mark-up |
|---|---|---|
| Exclusive control by store managers........... | 57.2% | 59.3% |
| Part control by store managers............... | 21.4 | 22.2 |
| Exclusive control by central headquarters...... | 21.4 | 18.5 |
|  | 100.0% | 100.0% |

Source: *Chain Stores: Chain-Store Price Policies*, p. 30. Price data for 28 chains; mark-up data for 27 chains.

## VI. Owner-Management Relationships

Another aspect of organization which may influence the process of price formation is the nature and location of control within the firm. In the context of owner-management relations the term control refers to the concentration of the power of decision in the hands of the owners of a small percentage of the firm's assets. It is to be expected that the policies of a department store controlled by a single owner will be different from those of a store controlled by a small corporate minority. Qualitative aspects of control are also important: direct owner participation in management is likely to give different results from those achieved by a professionalized management group. Banker control is likely to lead to different policies from those fostered by merchant control.[5]

Separation of ownership and control over a broad range of American industry has been studied by several investigators, and the results are now familiar to economists.[6] Unfortunately, although six of the largest non-financial corporations studied by the National Resources Committee in 1935 were engaged in the department-store business, no studies have yet been made of the extent to which ownership and control have become separated in this field. In 1927 it was observed that "by and large the management of the department

5. The comparison of financial and entrepreneurial types of control may over-simplify actual attitudes of management. Conservation of assets, for example, may lead a banker group to recommend expansion of output and reduction of prices — a reaction usually associated with operating management. Cf. Mason, "Price and Production Policies of Large-Scale Enterprise," loc. cit., p. 67.

6. Cf. A. A. Berle and G. C. Means, *The Modern Corporation and Private Property* (New York, 1932); *Structure of the American Economy*, Part I.

store is still in the hands of its owners."[7]   In those days, the owner was frequently the general manager, and, in smaller stores, merchandise manager as well.   "Where sits MacGregor, there is the head of the table;" organizational structure molded to the personality of the owner was ample evidence of the part played by MacGregor in policy-making.   In two decades several changes have taken place to reduce ownership control, and to place it partly in the hands of professionalized management, partly in the hands of a relatively small corporate minority.   In the first place, the striking growth of chains, already noted, suggests a steady decline in opportunities for owners to exert control over management.   Secondly a decrease in owners' active participation in management is probably indicated by the shrinking number of active proprietors of unincorporated department stores.[8]   Thirdly, as administrative subdivision has evolved under the impetus of growth, organization has become less plastic in the hands of individualistic merchant-entrepreneurs.   Even without changes in stock holdings, professional management has found ways of gaining effective control of the organization.   Finally, in listed corporations where there are large numbers of shares in the hands of the investing public[9] there is reason to believe that separation of ownership and control has followed the patterns prevailing in corporate organizations in other fields.

The chief result of these changes in owner-management relations appears to be greater reliance on the routinized functioning of the administrative machinery which has been described in earlier sections. Adjustments of the firm to competition come by planning and by rule: the budget of the merchandise manager and the controller triumphs over the shrewd insight of the individual merchant-entrepreneur.

## VII. Price Policies

We can now indicate certain price policies which are the logical outcome of the factors of internal organization already examined. It will be obvious that a complete explanation of the price policies of the department store must rest upon considerations of external market structure as well as those of internal organization; this further inquiry we leave for another occasion.   What follows, there-

---

7. Paul M. Mazur, *Principles of Organization* (New York, 1927), p. 31.
8. From 1441 in 1929 to 560 in 1939.   (*Retail Trade*, Part 1, p. 57).   This decline probably indicates that the owners of independent stores have sold out to chains and retired from active business life.
9. The three great chains, Sears, Ward and Penney, as well as most ownership groups such as Allied and Associated, are listed on the New York Stock Exchange.

fore, is in a sense a partial explanation of the policies discussed. Specifically, we examine the one-price policy, full cost pricing, price lining and leader selling.

The one-price policy, initiated by department stores in the post-Civil War period[1] and now almost universally followed in retail trade, calls for the sale of any given commodity to all consumers at the same retail price on any one day. In the absence of a one-price policy each retail transaction would be consummated only after sales clerk and customer had haggled over price in the manner of an Oriental bazaar. Such non-systematic price discrimination would almost certainly involve the store in great loss of good will on the part of consumers who had paid higher prices than their neighbors; and this fact was probably the chief consideration in the initial adoption of the one-price plan. Non-systematic price discrimination is also infeasible from the standpoint of department-store organization: waste of salesperson time, the need for selecting salespeople on the basis of bargaining ability, possibilities of dishonesty, and complex accounting problems which would result from non-systematic discrimination all indicate the need for a one-price policy. Without this policy it would be impossible to plan and control merchandising operations. Sales budgets, planned mark-downs and estimated gross margins would become meaningless in the face of unpredictable selling prices. It seems clear that the internal organizational processes of the store require the one-price policy.

Full cost pricing has been represented as the basic principle of pricing in retail trade.[2] This policy consists in adding to the wholesale cost of merchandise a certain percentage mark-up, set with reference to the estimated cost of doing an anticipated volume of business.[3] Included in the cost are overhead charges and normal

1. As early as 1653 Quaker merchants in England followed a one-price policy. This pricing principle obtained also in French department stores in the first half of the nineteenth century, the Bon Marché having adopted this policy sometime before 1838. In the United States, Rowland Macy traded at publicized fixed prices as early as 1854 when he was operating a store in Haverhill, Massachusetts. The one-price policy later became of central importance when he established his store in New York, R. H. Macy and Company. (Ralph M. Hower, *History of Macy's of New York, 1858–1919* (Cambridge, 1946), pp. 89, 26, 48–49).

2. The specific term "full cost" has been given currency in recent economic discussion by Hall and Hitch, op. cit. Its application to retail trade has been noted by many, (Chamberlin cites Mill, Cairnes and Wicksell, op. cit., pp. 106, 109). Recent examples are Due and Hawkins, cited above.

3. In passing, it may be noted that the logical necessity of simultaneous determination of mark-up and sales volume under the full cost policy results in circular reasoning on causes and results of price formation. Cf. Oliver, "Average Cost and Long-Run Elasticity of Demand," loc. cit.

profits (whatever seems appropriate in the light of past attainment and future expectations). Interest here attaches chiefly to the way in which budgetary procedures and administrative organization have led to the adoption and use of this policy. Once the expense rate for a department has been established on the basis of overhead charges and allocated arbitrarily among departments by accounting convention, mark-up can be set for the department as a whole. But, as was noted above, the mark-up rate is not in practice always that which will cover average cost thus computed. The term "full cost," therefore, is something of a misnomer; a more accurate designation of this policy would be an "average percentage mark-up" principle. We have already seen that the nature of the internal organization, especially the buyer-divisional manager relationship and the similar relationships between merchandise and control divisions, exerts pressure in the direction of bringing actual departmental average gross margin in line with the plan; although the "buyer" is permitted some liberty for dispersion of individual mark-ups about the departmental average. As it filters through the control organization of the department store, the so-called "full cost" pricing policy is reflected in a tendency for prices to hover about the level indicated by planned departmental mark-ups. This is a variation of the familiar theme: prices once set tend to remain fixed.

Price lining is the policy used by department stores of dividing their stocks of each type of goods into a limited number of price levels or groups. The practice of price lining is followed in lines where consumer demand, for reasons of taste or income, bulks large at certain modal points. It is particularly common in "shopping goods" lines where the consumer wishes to inspect an assortment of styles, colors, types and patterns before buying. Grouping all variations of a particular item into a small number of price lines simplifies the consumer's selection process and, by speeding sales, reduces selling costs. Price lining also permits the store to carry complete assortments in a limited number of price lines, at once satisfying the customer and reducing stock investment. While much of the incentive for adoption of price lining turns on the nature of certain department-store merchandise, the pressure to adopt this type of pricing comes also from the use of internal budgetary controls over the purchase and sale of physical units. The unit control system of a department store can operate efficiently only with a limited number of price lines. With unit control records, stocks are more easily managed and concentrated in a few important categories. Thus, buyers tend to "buy to" certain price lines, and their willingness to

purchase only goods which, at planned mark-ups, can be sold at such figures has led to the establishment in wholesale markets of corresponding practices. A department-store buyer may, therefore, be virtually compelled, by the absence of in-between prices in the wholesale market, to adopt certain price lines, and these, once established, tend to remain fixed for long periods of time. Effective price changes, therefore, often take the form of variations in the quality of the product offered at inflexible prices.

Leader selling is a price policy of cutting prices on certain items where elastic demand will draw large numbers of consumers into the store. Price reductions of this sort are intended to result in the *immediate* sale of goods other than the leader. The leader price policy is allied with the plan to sell other items at *normal* mark-up, increasing aggregate revenues. Use of leaders is dictated in part by imperfections of external market structure; thus Miller notes: "When the policy of employing leaders is adopted by a large retailer in the central market area which is surrounded by a series of interpenetrating marketing areas of lesser importance, the financial benefits to this dealer will be large by comparison with the losses of each of the surrounding dealers taken individually. If so, these latter may not follow suit since the basis of their sales appeal may lie rather in convenience, personal relations, or better services."[4]

Leader selling is also induced by internal factors of store organization. Departmentization permits the isolation of certain leader merchandise in a manner such that low mark-ups from leader selling will not adversely affect the operating performance of other goods. This is particularly desirable in the case of treating whole lines of merchandise as leaders. Work clothing, for example, may be placed in a single department and the entire line sold at low mark-up to induce traffic. Tea-room operations have frequently been conducted on the leader principle. Although logically it makes no difference whether the leader is handled in the same department with other goods or separated, in terms of actual store operation it is often the possibility of separate departmentization which makes this price policy seem feasible to the management, especially where the sales volume derived from the leader is substantial. Leaders, of course, may be individual items in a line rather than entire lines. Where this is true they will not be separately departmentized. Under these conditions pressure for the setting up of leaders may well come from the sales promotion division whose particular interests relate more

4. J. P. Miller, *Unfair Competition* (Harvard Studies in Monopoly and Competition No. 3; Cambridge, 1941), p. 253.

to the achievement of large sales volume than to the preservation of gross margin. Here again the internal structure of the store offers an explanation of the forces behind certain price policies.

## VIII. CONCLUSION

The proposition has been advanced that business men follow an imperfect marginalism. Certainly it is not impossible in retail trade, as in other lines of business, to produce evidence both of marginal economic conduct and of disregard of marginal calculations. Except in the rather meaningless sense that all conduct has some rational or irrational motivation (and therefore is directed toward a vague maximization), business behavior is not obviously guided by marginal principles. At any rate, another fruitful approach to the explanation of business phenomena such as price policies is through the examination of considerations internal and external to the firm which might shape the determination of these policies. In the case study of the department store set forth above we have studied several aspects of internal structure: the large size of the store, its administrative organization, its budgetary procedure, the interrelationship of units within the same corporate group, and the relations of owners to management. These factors encourage the use of such practices as: the one-price policy, full cost pricing, price lining and leader selling.

<div align="right">RICHARD M. ALT</div>

THE JOHNS HOPKINS UNIVERSITY

# THE MATHEMATICAL FOUNDATIONS
## OF ECONOMIC THEORY

SUMMARY

I. General comments: the use of mathematics by Hicks and by Samuelson, 111. — II. Samuelson's treatment of cost and production, 114. — III. Hicks and Samuelson on consumers' demand, 118. — IV. Samuelson's dynamics: difference equations, 121.

## I

There is no longer any doubt that mathematical methods are appropriately and usefully employed in the development of economic theory. The question, rather, is whether the mathematics should be discarded in the final exposition or whether they should take their place in the main argument. Do mathematics form the scaffolding or the steel framework of the structure?

Marshall used mathematical methods — relatively simple ones — as a scaffolding to assist him in constructing his theory; they were discarded when he had finished. The *Principles* suffer, I believe, from this fact; if we had been allowed to see more of the mathematical reasoning, we would have found fewer points of ambiguity and a generally tighter exposition. Be this as it may. The main points are that Marshall and many of his contemporaries were content with quite simple mathematical arguments and that the use of mathematics in economics has since developed both in scope and in complexity. The ways in which mathematics are used by many theorists are such that they cannot be discarded without leaving the argument defective and full of expressions such as "it can be proved that . . . ."

It is still possible, however, to confine the mathematical development to appendices. The completed structure can be described in general terms and, when the details of the construction need to be shown — when it is important to know that the structure has a steel frame — then reference can be made to technical appendices. This is the method adopted by J. R. Hicks, particularly in *Value and Capital*, which has appeared in a second edition incorporating important revisions and extensions.[1] It is, undoubtedly, a successful method in the hands of a master craftsman like Hicks.

The other possibility is the incorporation of mathematics into the

1. J. R. Hicks: *Value and Capital* (2d Edition, Oxford University Press, 1946).

main text, and this method is well illustrated by P. A. Samuelson, in his recently published *Foundations of Economic Analysis*.[2] Here the structure is built, stage by stage, from the foundations up. If the most difficult mathematics come early on, in the steel framework of the building, that's the way it is. The embellishments are not there to be admired until the frame is up. Samuelson's book may have less popular appeal than Hicks' but it is, none the less, of great importance. Every economic theorist worthy of the name will make a serious effort to examine it closely.

The two books of Hicks and Samuelson have much the same subject matter, the foundations of economic theory; they follow much the same principles, which I regard as undoubtedly the correct ones. They must be studied together, particularly in the light of the historical development of their respective theses. Since 1937 or 1938, each author has been re-shaping and rounding off his theory — which had taken a fairly comprehensive form before the war — and each has been influenced (though, unfortunately, rather remotely influenced in the geographical sense) by the work of the other. They have now come together on essentials. Future development, I believe, will not be Hicksian or Samuelsonian but will flow from an agreed combination of the two expositions. Postgraduate courses and seminars in economic theory will be concerned with this development for years to come.

Hicks and Samuelson, however, had different objects in mind in writing their books. Hicks tries to work out, if not a complete economic theory, at least a full development of one particular line of approach. He is not concerned if some of his conclusions can be reached by other and perhaps better roads. Samuelson's object is to unify diverse fields of economic theory by showing up the common, underlying mathematical basis. He is most concerned with how conclusions are reached, with what is valid and what is false in diverse theories.

Samuelson concentrates attention on operational or meaningful conclusions, i.e., conclusions which could, at least under ideal conditions, be validated or refuted by empirical data. His main unifying principle is that such conclusions are to be derived, not from the *equations* of equilibrium, but from the *inequalities* which ensure a maximum (minimum) position or which are required for stability. Another point he makes is that it is just as easy, and sometimes easier, to handle simultaneous change in many variables as in a few. The achievement of Walras and Pareto was to show the essential sim-

2. P. A. Samuelson: *Foundations of Economic Analysis* (Harvard University Press, 1947).

plicity of equilibrium of many variables. Samuelson's mathematical technique goes beyond Walras' and Pareto's and deals easily with many variable changes or shifts, something which is rather lost in Hicks' non-mathematical exposition of comparative statics.[3] A third point is that results can be obtained, and often easily obtained, in terms of finite changes and discontinuity, and not only in differential or continuous forms. Indeed, if conclusions are to measure up to empirical data, this is almost essential.

Like Hicks, then, Samuelson is much concerned with comparative statics, with the answers to such questions as: if demand shifts upwards, does the price increase? He goes on, moreover, to a sketch of what might be achieved in a field about which very little has been said, comparative dynamics. It is often thought that we progress naturally from statics to comparative statics and that we then switch into something different called dynamics, something better and more realistic. Samuelson emphasizes that statics and dynamics refer to the formulations of economic systems. Moreover, all formulations which involve time are not dynamic; a static system can easily include a secular or historical movement. Finally, Samuelson maintains, statics and dynamics cannot be kept as quite distinct branches of analysis. Comparative statics can be defined as the comparison of one position of equilibrium with another without reference to the path of transition from one to the other. This does not appear to involve any dynamic formulation. But Samuelson shows that meaningful conclusions in comparative statics come from conditions of stability of equilibrium positions. And these can only be derived from a dynamic model which shows under what circumstances a displacement from equilibrium will be followed by a return (perhaps oscillatory) to equilibrium.

It is just not possible in any one review to deal with all aspects of Samuelson's book. In the following sections, I concentrate on certain basic matters which happen to interest me. I would add that what makes the book such essential reading for the economist is not so much Samuelson's mathematical treatment as his economic insight. So many things which have worried the economist for years past appear so simple in Samuelson's devastating analysis. He is quite ruthless in casting out what is really irrelevant.[4]

3. Hicks' exposition is made possible by the use of the concept of a composite commodity on the assumption that the prices of the components vary in proportion. One result is that he gets rather tangled up with various concepts of "money."

4. This may be the place to register some (relatively minor) complaints. Samuelson has clearly been lax in editing and proofreading his book; there are numerous misprints and slips, some quite confusing, and the system of cross

## II

The most definitive parts of Samuelson's work are those on the theory of cost and production and on the theory of consumers' demand. Here he has undoubted success in his attempt to unify, simplify and codify existing theories. Many of the constructions and concepts which have exercised the minds of economists in the past are shown to be of little significance, except perhaps for expository purposes. The linear homogeneous production function, consumers' surplus, and the assumption of constant marginal utility of "money" are only some of the more notable casualties. Some may wish to revive them, but very potent restoratives will be needed. Samuelson himself has very little use for them; he remains somewhat uncertain only in the case of the "integrability" conditions on the consumers' scale of preferences, which (as an economist) he would like to dismiss but which (as a mathematician) he cannot quite ignore.

Samuelson differs from Hicks in his treatment of the theory of production. He takes it before, and not after, the theory of consumers' demand. He presents it in a stage by stage analysis, in contrast to the wider but more formal method of Hicks. There is room for both treatments although, personally, I prefer Samuelson's. His analysis throws more light on matters which have troubled economists and which have been the subject of heated controversy. The implications of "pure" competition, the "adding up" problem and the question of discontinuities in the production function are examples. Samuelson, however, assumes that the firm has only one product; he might well have indicated the obvious extensions to the case of joint production.

The stages in the analysis are: (1) the combination of inputs (at given prices to the firm) to minimize cost for a given output; (2) the choice of output to maximize net revenue to the firm; and (3) the external relations of a firm to the rest of the industry. The first

references is inadquate. He is sometimes obscure in his wording and he might well have spared us such monstrous concoctions as "monotonicity" (p. 12). He is not always happy on the mathematical side and tends to fall between two stools. His mathematical treatment is not simplified enough for the economist and not rigorous enough to satisfy the mathematician. (As a small example, on pp. 65–66, he says: "This can be proved rigorously in two ways . . .". But when he comes to the second proof he starts: "More rigorously . . ."). His compromise is particularly unsatisfactory in his handling of matrix algebra. He would seem to have decided to keep the matrix notation to footnotes, without detailed explanations; unfortunately, matrices have tended to creep back into the text in a rather untidy and confusing way. A text on matrix algebra suitable for the economist is badly needed and it is a pity that Samuelson has not added a third mathematical appendix to those in which he discusses quadratic forms and difference equations.

two can be run together, but the third raises different problems and must be kept separate.

With (1) as an illustration, I would make first a mathematical point which is not always appreciated. A problem of maximum (minimum) subject to single constraint can always be posed in two ways with identical results. Generally, the maximum (minimum) of one function is sought subject to the constancy of a second function. This can be turned round to give the minimum (maximum) of the second function subject to a constant value of the first. For example, let $x = x(v_1, v_2, \ldots \ldots v_n)$ be the production function and $y = y(v_1, v_2, \ldots \ldots v_n)$ be the cost function of a firm, the $v$'s being inputs. Stage (1), as posed above, is to minimize $y$ for given $x$. The solution is obtained by minimizing $(y - \lambda x)$ where $\lambda$ is a Lagrange multiplier, i.e.

$$\frac{\partial y}{\partial v_i} = \lambda \frac{\partial x}{\partial v_i} \qquad (i = 1, 2, \ldots \ldots n)$$

with          $x(v_1, v_2, \ldots \ldots v_n) = x = $ constant.

This gives $y$ and the $v$'s as functions of $x$ (and of the given input prices which are $\dfrac{\partial y}{\partial v_i}$). The alternative formulation is to maximize $x$ for a given $y$ (maximum output at a given cost.) Here we maximize $(x - \mu y)$, i.e.

$$\frac{\partial x}{\partial v_i} = \mu \frac{\partial y}{\partial v_i} \qquad (i = 1, 2, \ldots \ldots n)$$

with          $y(v_1, v_2, \ldots \ldots v_n) = y = $ constant.

giving $x$ and the $v$'s as functions of $y$ (and the input prices). The result is identical with the first, setting $\mu = \dfrac{1}{\lambda}$ and inverting $x$ as a function of $y$ into $y$ as a function of $x$.

In stages (1) and (2) together, with a continuous production function, one of the few meaningful results which can be obtained is $\dfrac{\partial v_i}{\partial w_i} < 0$ where $w_i$ is the price of the input $v_i$. The firm's use of an input decreases as the price of the input increases (other input prices constant). This is the continuous case. But little difficulty is found in the analysis, or in the interpretation of the operation of the firm, if certain discontinuities are assumed. Finite differences then replace differentials or partial derivatives, and the meaningful result appears as $\sum\limits_{i=1}^{n} \Delta v_i \, \Delta w_i \leqslant 0$. If all input prices are raised or

remain unchanged ($\Delta w_i \geqq 0$), then the amounts used cannot all increase. A difficulty with this kind of condition is to be sure about dropping or retaining the equality sign. (Samuelson is perhaps not always above criticism on this point.) Here the equality sign is essential. We cannot quite say that the amount of an input used will decrease if its price rises (other prices remaining unchanged). There is always the possibility, in the discontinuous case, that no change in the use of an input follows a rise in its price.

Samuelson has much that is sensible to say about the "adding up" problem, with particular reference to the linear homogeneous production function and the analysis of stage (3) when external forces are brought to bear on the firm. Many economists appear fascinated by the linear homogeneous production function, by the attainment of minimum unit cost under "pure" competition and by the complete distribution of the product with net revenue (or profits) zero. Here is complete confusion. For one thing, a U-shaped unit cost curve (i.e. a curve with a minimum at all) is not consistent with a linear homogeneous production function, for which unit and marginal costs are both constant. Samuelson attributes the confusion to a desire to explain everything in terms of marginal productivity with nothing left as a residual. Further, even if we want complete distribution, there is no need to demand it at all outputs (as with the linear homogeneous production function); it is sufficient to get it at one output (where unit cost is a minimum).

The first conclusion is that the assumption of a linear homogeneous production function should be dropped, except as a very special case which is "singular" in the mathematical as well as the general sense. If we feel we need some explanation of why doubling the use of all factors does not double the output, we might agree with Samuelson to include in the production function, not all "factors," but only "inputs" limited to measurable economic goods and services. Or we might quote Hicks' assumption that the firm "possesses some fixed productive opportunity which limits the scale of production, and to which the surplus can be imputed as earnings."[5] Or we might follow Chamberlin and maintain that, while, with larger input aggregates, increased specialization and technical economies may dominate at first, eventually they will be more than offset by rising costs of co-ordination.[6]

5. Op. cit., p. 322.
6. E. H. Chamberlin, "Proportionality, Divisibility and Economies of Scale," this *Journal*, February 1948. Chamberlin appears to consider the more general case where all factors are variable and included in the production function, as against Samuelson who excludes, and Hicks who fixes, some factors. I doubt,

The production function not being homogeneous, the unit cost curve can take the familiar U-shape. Under "pure" competition, the price of the product being given to the firm, equilibrium for stage (1) and (2) combined occurs at the output which makes marginal cost equal to price. This is not at the point of minimum unit cost, except by accident, and the firm makes a profit. (The profit could be a loss but then the firm goes out of business in the long run.) The common objection is that this is inconsistent with "pure" competition. The concept of "pure" competition here is clearly wider than the one used above (where the prices of product and of inputs are given to the firm). Something new has been added, something of a quite different nature, namely the condition that the firm's profit be zero, implying that unit cost is minimal. It is not always clear which of these two, zero profit or minimum unit cost, is to be regarded as the condition. But, whichever it is, it is an inappropriate addition which serves only to confuse the issue.[7]

Having got the equilibrium of the firm under "pure" competition (in the above sense) or under other conditions, we can indeed proceed to a quite distinct analysis, that of stage (3). Here we seek to explain

however, whether he differs from the others, except in exposition. Would any of them object to the following?

The theory of the firm deals with an entity which, whatever adjustments it makes in its organization, remains the co-ordinating and decision-making unit. The production function is defined for, and is peculiar to, this entity. All factors which can be measured in quantitative units appear as variables in the function. (If a "factor" is not so measurable, how can its use be doubled?) The function expresses the maximum output the firm can get from stated amounts of the factors, the best combination of the factors the firm can achieve as a co-ordinating unit. The form of the function reflects the firm's co-ordinating power in its own environment; it shows what the firm is technically capable of achieving. Not all aspects of co-ordination appear in the form of the function; some are expressed as measurable factors (such as the use of punch-card equipment). But there is always something — some non-measurable "factor" — which can only show up in the form of the function. The net revenue the firm can derive, as a maximized residual or surplus, is dependent on the technical conditions expressed in the production function in conjunction with the conditions of factor supply and of demand for the product. The net revenue of one firm can differ from another's, not only because it faces different supply-demand conditions, but also because its production function is different. As indicated here, the determination of maximum net revenue can be shown in two stages. The first fixes the proportions in which factors are combined for any scale of production; the second determines the scale of production. Chamberlin's analysis now falls into place. And at no point do we feel driven to assume a linear homogeneous production function, except as a singular case.

[See below, p. 143, note 7, for comment — Ed.]

7. It can be noticed that "pure" competition was introduced (Chamberlin: *The Theory of Monopolistic Competition,* pp. 6–7) precisely because the concept of "perfect" competition had grown too wide and variable. It has since tended to suffer the same fate.

how the level of profits is determined in any industry. In particular, we can lay down certain conditions on "free entry," on potential competition from outside firms. The conditions may imply — though this is not necessary — that net revenue for the marginal firm is zero, and hence that this firm's output is at minimum unit cost. As Samuelson points out, however, there is no empirical evidence in real life that there are "competitive" forces driving profits to zero. But there are many ways in which "free entry" conditions can be framed; the point here is that they are additional to the analysis of the firm. They can be added, or left to be swallowed up in the wider analysis of market equilibrium where prices of products and inputs are determined.

## III

In the theory of consumers' demand, we have a single consumer with given income $I$ able to purchase $n$ goods on the market at given prices $p_1, p_2, \ldots \ldots p_n$. We want, first, to be able to say that the consumer's demand is uniquely determined, i.e., that his purchases, $x_1, x_2, \ldots \ldots x_n$ are single-valued functions of $I, p_1, p_2, \ldots \ldots p_n$. The demand, moreover, should be invariant to proportionate changes in income and all prices, i.e., the demand functions are homogeneous of order zero. Secondly, as a problem of comparative statics, we want to trace how demand changes as income and prices are varied. In particular, we seek general restrictions on these changes, restrictions which (ideally) could be verified or refuted from empirical data.

We must expect that there are several constructions, or sets of hypotheses, which serve to "explain" the same phenomena. Our choice must then be for the simplest model. As economists, we do not wish to introduce complications not needed for an explanation of empirical economic facts. In particular, we have no direct interest in features of consumer behavior which may appeal, for example, to the psychologist.

One explanation of consumers' demand is based on the assumption of an ordinal preference field with properties ensuring that, for any given income and prices, there is a unique set of purchases at the maximum level on the preference scale. Hicks has shown how neatly this can be worked out. There are certain *equations*, invariant with respect to an ordinal utility function, which determine equilibrium and show how the demand functions of the required form arise. There are certain *inequalities*, again invariant with respect to the ordinal utility function, which ensure a maximum rather than a minimum. These can be translated into restrictions on the variation of the

demand functions as income and prices change. As in the theory of production, these conditions can be obtained from alternative formulations of the problem; the consumer *either* maximizes utility for a given income *or* minimizes expenditure for a given utility level, prices being given in each case. We can choose between the formulations according to convenience of exposition.

In developing his theory, Hicks was disturbed by the fact that he could not use all his inequalities in deriving restrictions on the demand functions. Samuelson has now shown, what Hicks had begun to suspect,[8] that there is just one inequality, and a very simple one, from which we can derive all restrictions on demand. The inequality turns on "cross" valuations of purchases at different sets of prices in the manner of standard index number practice. Indeed, the inequality serves as the basis, not only for the analysis of demand functions, but also for index number theory and other purposes.[9]

We compare two equilibrium situations, denoted by suffixes 0 and 1, for the consumer. Writing the summation $\Sigma$ over the $n$ goods, we first compare the cost of the purchases of the second situation at the prices of the first with the actual cost of the purchases of the first situation, i.e., we compare $\Sigma p_0 x_1$ with $\Sigma p_0 x_0$. If $\Sigma p_0 x_1 \leqslant \Sigma p_0 x_0$, then the quantities $x_1$ could have been purchased in the first situation with the income then available ($\Sigma p_0 x_0$), whereas in fact the quantities $x_0$ were purchased. The consumer prefers the quantities $x_0$ to the quantities $x_1$. The second comparison is between $\Sigma p_1 x_0$ and $\Sigma p_1 x_1$. Here, if $\Sigma p_1 x_0 \leqslant \Sigma p_1 x_1$, then the consumer prefers the $x_1$ quantities to the $x_0$ quantities. For consistent consumer behavior, if the first inequality holds, the second cannot hold, and conversely. Hence, $\Sigma p_0 x_1 \leqslant \Sigma p_0 x_0$ implies $\Sigma p_1 x_0 > \Sigma p_1 x_1$. Translating into finite difference form, we have

$$\text{if } \Sigma p \Delta x \leqslant 0, \quad \text{then } \Sigma (p + \Delta p)\Delta x < 0 \quad \ldots \ldots \quad (1)$$

where $\Delta p$, $\Delta x$ are changes from one situation to the other. (1) is the inequality condition which Samuelson shows to be basic to the theory.

Now, the last paragraph, leading to (1), does not refer at all to a utility function or a preference scale, only to what Samuelson calls "revealed preferences." If we simply lay down (1) as our basic hypothesis — and it is in a form (ideally) refutable from empirical data — then we get practically everything we want. First, it follows that the demand functions are unique and homogeneous of zero

8. Op. cit., pp. 51–2, 311, and Additional Note A.
9. Cf. J. R. Hicks: "The Valuation of the Social Income," *Economica* (1940), and "Consumers' Surplus and Index Numbers," *Review of Economic Studies* (1942).

order.[1] Secondly, a particular case included in (1) is

$$\text{if } \Sigma p\Delta x = 0, \quad \text{then } \Sigma \Delta p \Delta x < 0 \quad \ldots\ldots\ldots \quad (2)$$

The meaning of this needs to be examined carefully. There have been changes ($\Delta p$) in the prices. There has also been a change in the income of the consumer:

$$\Delta I = \Sigma(p + \Delta p)(x + \Delta x) - \Sigma px = \Sigma x \Delta p$$

i.e., the price changes have been compensated by a change $\Sigma x \Delta p$ in income. The compensation is such that the quantities now purchased (following the compensated price changes) could have been bought at the old prices and income: $\Sigma(x + \Delta x)p = \Sigma xp$. For such compensated price changes, $\Sigma \Delta p \Delta x < 0$. If the price changes are all upwards (or some up and others unchanged), then we cannot say that all demands are decreased but we can say that on balance there is a decrease in demand. In particular, if there is a rise ($\Delta p$) in one price only (other prices unchanged) compensated by an increase ($x\Delta p$) in income, then the demand for the good falls.[2] Finally, in the limiting form with differentials and rates of change, (2) becomes $\Sigma dp\, dx < 0$ if $\Sigma p\, dx = 0$ so that $dI = \Sigma x\, dp$ is the compensating change in income. Hence

$$\sum_{r=1}^{n} \sum_{s=1}^{n} \left( \frac{\partial x_s}{\partial p_r} + x_r \frac{\partial x_s}{\partial I} \right) dp_r dp_s < 0 \quad \ldots\ldots\ldots \quad (3)$$

We are now on familiar ground. From (3), if only $p_r$ changes, then

$$\frac{\partial x_r}{\partial p_r} + x_r \frac{\partial x_r}{\partial I} < 0$$

which is Hicks' substitution effect. On the other hand,

$$\left( \frac{\partial x_s}{\partial p_r} + x_r \frac{\partial x_s}{\partial I} \right)$$

can be of either sign, which is Hicks' complementarity. These expressions are, of course, Slutsky's compensated variations in demand.

It appears, therefore, that all we need about demand derives from the single condition (1). There is no need to postulate an

1. Samuelson, op. cit., pp. 111–12.
2. If we admit the hypothesis of an ordered preference field, then there is another compensated price change, namely a change in prices accompanied by a change in income sufficient to maintain the consumer on the same indifference surface. The compensated price change of the text maintains the consumer on the same price plane, not the same indifference surface. The two compensations are thus different for finite changes; they coincide for differential changes. It can be shown, however, the $\Sigma \Delta p \Delta x < 0$ holds equally for both. Samuelson does not make the distinction here drawn.

indifference map or to assume that the consumer acts to maximize utility or minimize expenditure. We gain a good deal by concentrating on the condition (1) and shaking ourselves free of traditional utility analysis. For one thing, we get away from the concept of elasticity of demand with respect to variation in a single price; it is really easier to let all prices vary together. Again, we find we can operate with finite changes (and not only with rates of change) and so approach nearer to empirical tests. To take one example, we can follow Hicks[3] in extending (1) from a single consumer to the whole body of consumers. If we find $\Sigma p_0 x_1 < \Sigma p_0 x_0$ and $\Sigma p_1 x_0 > \Sigma p_1 x_1$ are both true, then (1) is satisfied. We know that the behavior of consumers is consistent in the two periods and we can say that there has been a fall in real social income. On the other hand, we could find $\Sigma p_0 x_1 < \Sigma p_0 x_0$ and $\Sigma p_1 x_0 < \Sigma p_1 x_1$ which contradicts (1); something has happened between the two periods to change the pattern of consumers' behavior. This is an exercise in the New Welfare Economics.

But does it follow that we should scrap utility analysis altogether? We should not be too hasty. I am inclined to think that we should keep utility analysis, as developed by Hicks and others, at least for purposes of exposition and interpretation. It is useful to be able to say, when faced with the contradiction just mentioned, that it seems probable that the indifference maps of consumers have changed. It is useful to be able to describe the well-known terms in (3) as compensated price changes, corresponding to the maintenance of a given utility level. Any change whatever in prices is conveniently split into a change along the indifference surface (the substitution effect) and a proportionate change in all prices (the income effect).

## IV

So far, equilibrium positions have derived from conditions of maximum (minimum). The inequalities which ensure a maximum rather than a minimum (or conversely) then give the operative results in comparative statics. The mathematics of the analysis turn largely on the theory of quadratic forms. The inequalities are the stability conditions and they do not involve, except implicitly, considerations which can be described as dynamic.

In other problems, equilibrium positions arise, not from conditions of maximum (minimum), but from the equation of certain variables. Questions of comparative statics now involve stability conditions which are more difficult to formulate and which (as Samuel-

3. "The Valuation of the Social Income," *Economica* (1940).

son shows) must be based explicitly on dynamic models. A typical case is the equation of demand and supply on a market, and stability considerations can lead to the familiar "cobweb" theorem. Dynamic systems are needed in comparative statics just as much as in (say) a theory of industrial fluctuations.

The definition of dynamics adopted by Samuelson follows Frisch. A system is dynamic "if its behavior over time is determined by functional equations in which 'variables at different points of time' are involved in an 'essential' way."[4] From the mathematician's point of view, therefore, dynamic relations are essentially difference equations. It may be appropriate sometimes to introduce continuity, to replace finite differences by rates of change, and sums by integrals. The relations are then differential, integral or mixed equations. But, whereas in physics differential or integral equations are usually appropriate, this seems to be less often the case in economics. A ball set rolling down a hill will continue to roll until it reaches the bottom. However, even in the most rapid collapse of prices on the stock exchange, prices are only marked down at definite points of time and business is suspended at closing time.

Economic dynamics must thus be based firmly on a theory of difference equations. Hitherto no account of such a theory has been available to economists. Samuelson makes good the deficiency in a long mathematical appendix. But it is very tough going, partly because the exposition is formal and lacks illustrative examples designed to maintain the economist's interest. Examples are useful, not only in illustrating general methods, but also in showing how actual difference equations arise in economic problems. The following examples are offered in the hope that they may help economists to follow Samuelson's formal development.

A difference equation involves variables at different points of time. Perhaps the simplest example is the growth of a sum at interest, of $100r$ per cent per year compounded annually. If $y(t)$ is the amount at the end of year $t$, then

$$y(t + 1) = (1 + r)\, y(t) \qquad \ldots \ldots \ldots (4)$$

This relation involves two points of time or one time lag; it is a difference equation of the first order. It is also linear since only $y(t)$ and $y(t + 1)$ appear, and not powers, etc. Finally, it is non-historical since time does not appear explicitly; the relation is not dated but holds for any two consecutive years whatever. A general solution of (4) is an expression of $y$ as a function of $t$, and it can be written

4. Samuelson, op. cit., p. 314.

once a single initial condition is specified, a condition which serves to "start up" the system. Suppose we specify the initial sum $y_0$. Then we have the following step-by-step solution:

$$y(0) = y_0; y(1) = (1 + r)y(0) = y_0(1 + r);$$
$$y(2) = (1 + r)y(1) = y_0(1 + r)^2; \ldots \ldots$$

and generally $\quad\quad y(t) = y_0(1 + r)^t$

A slightly more complicated case arises when the difference equation has a term not involving the variable $y$. For example, in the compound interest problem, suppose that a fixed bonus of $a$ is added each year. Then $y(t + 1)$ is the sum of $(1 + r)y(t)$ and $a$, i.e., the difference equation is

$$y(t + 1) - (1 + r)y(t) = a \quad\quad\quad \ldots\ldots\ldots (5)$$

With the same initial condition, the step-by-step solution is now

$$y(t) = y_0(1 + r)^t + a(1 + r)^{t-1} + a(1 + r)^{t-2} + \ldots\ldots + a$$

or $\quad\quad y(t) = y_0(1 + r)^t + \frac{a}{r}[(1 + r)^t - 1]$

The general solution of (5) is in two parts. One part is $y_0(1 + r)^t$ which is the general solution of the "reduced equation" $y(t + 1)$ $- (1 + r)y(t) = 0$. The other part is $\frac{a}{r}[(1 + r)^t - 1]$ which can be checked to be a particular solution of the full equation (5).[5] This suggests what is indeed one of the main working rules for solving difference equations. To solve (5), we first find by trial and error some particular solution of it, and we then add the solution of the simpler equation (4) to get the general solution.

A further complication arises when the difference equation is historical, involving time explicitly. Suppose that a bonus is added each year as in (5) but that its amount is declared year by year. Let 1929 be the initial year ($t = 0$) when the sum $y_0$ is put at compound interest. Suppose, as a matter of history, the bonuses declared are $a_1$ in 1930 ($t = 1$), $a_2$ in 1931 ($t = 2$), $\ldots\ldots$ Then

$$y(t + 1) - (1 + r)y(t) = a_{t+1} \quad\quad\quad \ldots\ldots\ldots (6)$$

is the difference equation. The solution is found step by step to be

$$y(t) = y_0(1 + r)^t + a_1(1 + r)^{t-1} + a_2(1 + r)^{t-2} + \ldots\ldots + a_t$$

---

5. Put $y(t) = \frac{a}{r}[(1 + r)^t - 1], y(t + 1) = \frac{a}{r}[(1 + r)^{t+1} - 1];$ then

$$y(t + 1) - (1 + r)y(t) = \frac{a}{r}[(1 + r)^{t+1} - 1] - \frac{a}{r}[(1 + r)^{t+1} - (1 + r)] = a$$

The solution of (6) is the sum of the solution of the same "reduced equation" and a particular solution which depends on the historical run of bonuses.

We can now pass to a difference equation of second (or higher) order involving two (or more) time lags. The well-known auto-regressive system of Yule and Kendall provides an illustration.[6]

The second order case, in linear form, is:

$$y(t + 2) + ay(t + 1) + by(t) = \epsilon_{t+2} \quad \dots\dots\dots (7)$$

where $a$ and $b$ are constants of the structure of the system, and where $\epsilon_{t+2}$ is an historical disturbance. Any term in the series is determined by the two preceding terms, except that a disturbance peculiar to the term in question is added. Without the disturbance, we have the non-historical form:

$$y(t + 2) + ay(t + 1) + by(t) = 0 \quad \dots\dots\dots (8)$$

Here (8) is the "reduced equation" corresponding to (7). The method of attack is as before. We find a general solution of the reduced equation (8), a solution which involves only the structural constants $a$ and $b$ but which needs two initial conditions, e.g., $y(0) = y_0$ and $y(1) = y_1$, to "start up" the series. Next, we try to get some particular solution of the full equation (7). The sum of the two provides the general solution of (7). The general solution, then, consists of a systematic part independent of the historical disturbance and a part which arises historically. What we look for in the first part generally is a damped oscillation. This kind of system is clearly relevant to certain theories of industrial fluctuations.

The general method is very similar to that adopted in the theory of differential equations. There are some rather formal "existence" theorems which ensure that a unique solution exists.[7] In practice, success depends on trying out a box of tricks, e.g., to find some particular solution of (7) to add to a general solution of (8). It is all rather scrappy and hit-or-miss. With difference equations, however, we can always fall back on the step-by-step method to trace the course of the variables, given by specific difference equations, starting from specified initial conditions.

The use of "operators" is very helpful. Let $E$ be an operator which turns $y(t)$ into $y(t + 1)$, i.e., $Ey(t) = y(t + 1)$. Then $E^2y(t)$

---

6. See M. G. Kendall: *Contributions to the Study of Oscillatory Time Series* (Cambridge, England, 1946).

7. The existence theorems for difference equations are rather glossed over by Samuelson, e.g., op. cit., p. 385.

$= Ey(t + 1) = y(t + 2)$, and so on. Let $\Delta$ be the usual operator for finite differences. Then $\Delta y(t) = y(t + 1) - y(t)$; $\Delta^2 y(t) = \Delta y(t + 1) - \Delta y(t) = y(t + 2) - 2y(t + 1) + y(t)$; and so on. Any difference equation can be represented alternatively in terms of $E$ and in terms of $\Delta$. Equation (4) can be written

$$[E - (1 + r)]y(t) = 0 \quad \text{or} \quad (\Delta - r)y(t) = 0 \quad \ldots \ldots \ldots \quad (9)$$

Equation (8) can be written

$$(E^2 + aE + b)y(t) = 0 \text{ or}$$
$$[\Delta^2 + (a + 2)\Delta + (a + b + 1)]\, y(t) = 0 \quad \ldots \ldots \ldots \ldots \ldots \quad (10)$$

To get the second form, some rearrangement of terms is needed in each case. But the utility of operators in practice arises from the fact that they can be handled as though they were quantities. By definition

$$\Delta y(t) = (E - 1)y(t)$$

We can write $\Delta = E - 1$ or $E = \Delta + 1$ and transform the equation in one form to the other. So, in (10)

$$E^2 + aE + b = (\Delta + 1)^2 + a(\Delta + 1) + b$$
$$= \Delta^2 + (a + 2)\Delta + (a + b + 1)$$

It is to be noticed that a difference equation written in terms of finite differences (with the operator $\Delta$) is in analogous form to a differential equation in which the operator is $D = \dfrac{d}{dt}$. For example, the difference equation (9) for interest compounded annually is $\Delta y(t) = ry(t)$. The corresponding differential equation for interest compounded continuously is $Dy(t) = ry(t)$.

The nature of difference equations which arise in the dynamic formulation of a problem in comparative statics can be illustrated by a simplified example used by Samuelson. Equilibrium on the market for a single good is determined by the equation of demand and supply: the equilibrium price $\bar{p}$ is a solution of $D(p) = S(p)$. If price is initially fixed at $p_0 (\neq \bar{p})$, what happens over time? Of the many different dynamic formulations of this problem, two can be noticed here. First, we start with the assumption that price will rise as long as demand exceeds supply and will fall as long as supply exceeds demand. We add an assumption, which involves continuous changes, that the speed of adjustment is proportional to the difference between demand and supply. A differential equation is the appropriate expression of the dynamic relation:

$$\frac{dp}{dt} = \lambda[D(p) - S(p)] = \lambda[D'(\bar{p}) - S'(\bar{p})]\,(p - \bar{p})$$

approximately by expansion. Here $\lambda$ is the speed of adjustment and positive. The equation is easily solved to show that $p$ moves continuously towards or away from $\bar{p}$ according as $S'(\bar{p})$ is greater or less than $D'(\bar{p})$. It is a feature of differential equations (such as this) that they permit only continuous movement one way or the other and exclude the possibility of oscillatory movements. For this reason alone, they are not altogether appropriate in economic dynamics.

Secondly, we start by assuming that demand adjusts itself immediately to price but that supply reacts only after a time lag. Supply at time $(t + 1)$ is $S[p(t)]$ when $p(t)$ is the price at time $t$. Demand at time $(t + 1)$ is $D[p(t + 1)]$ and $p(t + 1)$ is fixed to clear supply:

$$D[p(t + 1)] = S[p(t)]$$

This is a difference equation which, from a given initial price $p_0$, gives step by step the price $p(t)$ as it develops over time. In the particular case of linear demand and supply (with constants $\alpha$, $\beta$, $a$ and $b$), we have

$$\alpha - \beta p(t + 1) = a + b p(t)$$

i.e.,
$$p(t + 1) + \frac{b}{\beta}\, p(t) = \frac{\alpha - a}{\beta}$$

which is a linear difference equation of the same type as (5). The solution, easily derived, shows a greater range of possibilities; in particular, $p$ can oscillate around, towards or away from $\bar{p}$. This is the familiar "cobweb" result.

This example indicates the limitation of Samuelson's work. Why should supply at time $(t + 1)$ depend on price at time $t$? We might say that it depends on what suppliers at time $t$ *expected* the price at time $(t + 1)$ was going to be, and that the expectation depends on the price at time $t$. But the expectation depends on other things as well. The concept of expectation is one which finds no prominent place in Samuelson's theory. He gives a comprehensive and satisfying treatment of comparative statics, including the dynamic formulation of stability conditions. But he does not provide anything like the Hicksian analysis based on expectation. On economic dynamics in the wider sense, and particularly on trade cycle theory, he is content with some indications of what is to be done, making valuable suggestions of interest particularly to the mathematician. He points to the need to explore the almost unknown territory of comparative dynamics in which the whole course of an economic system over time can be varied by some change (e.g., in

the marginal propensity to consume). He stresses the potential importance of non-linear difference or differential equations and of systems with stochastic elements. We are here in a mathematical domain presenting formidable problems and still awaiting systematic development.

Mathematical economics in the past has been dominated by the mathematical convenience of linear systems. It seems likely that linear assumptions are not adequate in the treatment of economic dynamics. If so, some other simplification is needed to prevent the mathematics from getting too formal. No one has yet hit upon the appropriate simplification. There is much to be done by the mathematician as well as by the economist.

R. G. D. ALLEN.

LONDON SCHOOL OF ECONOMICS

# PROPORTIONALITY, DIVISIBILITY, AND ECONOMIES OF SCALE: TWO COMMENTS

Professor Chamberlin describes the subject of his article as "the long-run average cost curve of the firm, interpreted as the *joint* result of the proportions of factors employed and of their aggregate amount."[1] He criticizes the practice of treating proportions and size as separate problems, and asserts that "the erroneous thesis has come to be widely held that under the 'perfect divisibility' of theory, as applied to the factors of production, there would be no economies or diseconomies of scale." In discussing these matters, it would seem to be advisable, first, to distinguish carefully between the situation in the real working world and that in any particular theoretical model, and, second, to define rather carefully what is meant by "divisibility." Professor Chamberlin acknowledges the importance of the second requirement,[2] although in fact he does not attempt an explicit definition; but he says nothing about the first. His article therefore suffers from a double confusion: (1) it is not clear whether he is speaking of the real world or of some particular theoretical model, and (2) the reader must decide for himself, from the context, what is meant by "divisibility."

It is customary to explain economies of scale by the "lumpiness" of certain factors of production. For various technical reasons the factor may be available only in discrete units, often of substantial size, and therefore over a considerable range output may be variable only by altering the amount of other factors used in conjunction with the "lumpy" factor. From this point of view, the problem can be said to involve the *proportions* in which the factors are combined; there is a certain optimum proportion of factors, but this optimum is only attainable when the aggregate use of the lumpy factor is large,[3] since it can be had only in (large) discrete units. Imperfect divisibility, in the sense that it is impossible to obtain the lumpy factor in fractional units having proportionate efficiency,[4] makes it impossible to produce with equal efficiency at lower outputs.

1. This *Journal*, LXII (February 1948), p. 229.
2. Ibid., p. 239.
3. "Large" in a relative sense only. Logically the argument applies equally to one among many local barber shops and to a single steel mill capable, at optimum size, of supplying an entire economy.
4. In practice this may be thought of as inability to get fractional units bearing the same ratio of cost to output as standard units.

Professor Chamberlin objects to this approach. He rejects any definition of divisibility that requires or implies that the efficiency of the factor is not affected,[5] though without offering any reason except that such a definition makes a mere tautology of the proposition that economies of scale depend on imperfect divisibility. But the mere fact that a proposition is tautological does not necessarily make it useless or unimportant. Professor Chamberlin's argument may also be dismissed as tautological once his (implicit) definition of divisibility is granted; indeed, some philosophers have contended that all reasoning can be reduced to a series of tautologies. In effect he simply defines his terms to include only those types of divisibility that are practicable in the real world, and he discusses no other possibility. It is not surprising, therefore, that he comes to a conclusion that appears to be very different from that of other economists; although he uses the same terms as they, he imbues them with an entirely different meaning. Having defined his terms so as to exclude the influence of indivisibility in the sense used by others, he finds that indivisibility in this sense plays no part in his model. What could be more tautological than that? Thus, as is so often the case in economics, the crucial factor is the choice of definitions and assumptions: once these are settled, the rest is hardly more than an exercise in logic, and there can be little room for argument about the conclusions to be reached.

That it is tautological does not necessarily invalidate Professor Chamberlin's proposition, however, any more than it does that of those whom he criticizes. It becomes, rather, a question of whether his definitions are more useful for certain purposes than the other definitions — in other words, whether they more accurately reflect certain aspects of the real world. So far as the practical application of economic theory is concerned, it is certainly more realistic to think in terms of models in which factors of production are subject to the limited types of divisibility which Professor Chamberlin discusses. As one who is primarily interested in economics for its practical application to the world's problems, rather than as an arm-chair occupation for the intellect, this correspondent would have heartily agreed, had the approach been from such a viewpoint. But Professor Chamberlin goes far beyond that, and seems to believe that he has disproven the "indivisibility" thesis in any and all circumstances. Yet the broader definition of indivisibility, which includes the qualification that efficiency is not to be affected, is every bit as legitimate as his narrower definition, and is certainly useful in understanding the forces at work. It is quite true that in the real world

5. Ibid. pp. 237 ff.

the long run average cost of the firm is "the *joint* result of the proportions of factors employed and of their aggregate amount." It is not *wrong*, however, to assert that there would be no economies of scale *if* all factors were perfectly divisible (as broadly defined). It would be very unwise to apply uncritically the conclusions derived from such a model, but there is nothing inherently wrong in the model itself; and when properly used it is helpful in understanding the real world.

The difference between Professor Chamberlin and those whom he criticizes comes to no more than this: they are concerned with the fact that in certain circumstances it is impossible to apply certain factors of production with equal efficiency except in discrete units, whereas he is concerned with the fact that in these same circumstances any attempt to apply these same factors other than in discrete units (e.g., by resort to part-time practices) involves some loss of efficiency. There is evidently no disagreement about the facts, only a difference in approach and terminology. Yet surely the broader definition of indivisibility is more in accord with common usage. Let Professor Chamberlin complain, if he will, when someone speaks of production in the real world as if it were merely a question of proportions of factors without regard to their amount; that has little or nothing to do with the fact that, if all factors of production were capable of employment in infinitesimal increments with exactly proportionate efficiency (a condition that no one contends is met in the real world), there would be no economies of scale, and proportionality of factors would in fact be the only significant consideration. It is certainly true that a skilled laborer can not be divided up with a meat cleaver and apportioned among several employers without loss of productive efficiency,[6] but it is equally certainly true that this very indivisibility or lumpiness of factors is what makes economies of scale possible.

It may also be pointed out that Professor Chamberlin's refusal to discuss increments in the amount of a factor used by a given firm on the assumption that the increments are of unvarying unit efficiency is at variance with the usual practice in economic analysis. When describing the changes in total output that occur when additional units of labor are applied to a given fixed plant, for example, it is customary to stipulate that all units are conceived to be of equal efficiency. Everyone knows that in the real world a firm does not and cannot hire labor units of standard efficiency, but a number of individual men, who vary greatly in ability, in strength, in energy, and in numerous other relevant ways; yet the concept of labor units of standard efficiency is highly useful in deriving the underlying

6. Cf. Ibid., p. 240.

economic relationships, unencumbered by the effects of variations in the abilities of individual laborers. It is obvious that, in any application of the theory to the real world, account must be taken of the fact that laborers are actually individuals and not standard units of labor; a monograph on the adjustments necessary for a practical application of the theory would be highly useful and valuable, but it could not be construed as a disproof of the theory. From all of which we may draw two conclusions with respect to Professor Chamberlin's arguments: (1) the customary approach (for whatever it is worth) favors the definition of divisibility which he rejects; and (2), more important, we are led back to the proposition already made, that his arguments are more appropriate to a description of divisibility in the real world (which he does not attempt) than to a criticism of divisibility as it enters certain theoretical models propounded by Knight, Stigler, Kaldor, Lerner, Boulding, and others (which he does attempt). Professor Chamberlin has missed an excellent opportunity of bridging the gap between theory and practice in this field.

<div style="text-align:right">A. N. McLeod</div>

WASHINGTON, D. C.

In a recent article in this *Journal*, Professor Chamberlin has attacked the proposition that economies or diseconomies of scale can occur only if at least one of the factors used in production is either fixed or indivisible. He contends that the above formulation is either tautological or simply wrong, and that economies and diseconomies of scale can and will occur in the absence of indivisibilities.

Professor Chamberlin treats the proposition that perfect divisibility and constant returns to scale are synonymous with great scorn, on the ground that the definition of divisibility used by most authors who maintain the above theorem already assumes what it is intended to prove. I venture to submit (1) that the question of a "proof" does not arise in this discussion, the whole problem being one of definition, and (2) that the only definition feasible is one which identifies perfect divisibility with constant returns.

The fact that most authors with whom Professor Chamberlin finds fault in this matter emphasize the tautological nature of their formulation, is acknowledged by him (although sometimes he appears to be implying that they intended something more). Such being the case, I take it that his major objection turns on the question of the usefulness of the tautological definition used. It is this problem to which we must now turn.

## I

It is surprising that throughout the discussion Professor Chamberlin never touches on the difficulties encountered in defining a "factor of production." Now, as far as the firm is concerned, the only satisfactory definition of a factor of production is one based on efficiency. If we are to talk of demand and supply curves of factors of production to the firm, we must define a factor in a manner such that the marginal rate of substitution between its individual units is equal to unity for all levels of output.

Thus we can measure any input in arbitrary units "$a$", and classify all those units between which the marginal rate of substitution is unity into the category $Fi$. If we say that any unit "$a$" within that category is divisible, this must mean that membership in the category is independent of the size of the input, where size is measured as a fraction or multiple of the standard unit "$a$". This being so, it is clear that, if input measured in $ka$ ($k < 1$) units is to continue as member of the category, then (1) the marginal rate of substitution between an amount $ka$ and another similar amount must be equal to unity, and (2) the marginal rate of substitution between $1/k$ the amount $ka$ and $a$ must also be equal to unity.[1]

It should be noted that the second postulate does *not* entail the assumption that an amount $ka$ is $k$ times as efficient as an amount $a$. It simply means that we can substitute $1/k$ units $ka$ for one unit $a$ without loss in efficiency. This is, of course, perfectly compatible with either increasing, constant or diminishing returns to successive amounts $ka$ of the category $Fi$.

Now let us suppose that a commodity $A$ is manufactured by the use of two categories of input, $Fi$ and $Fj$. Let the standard unit of $Fi$ equal $a$, and that of $Fj$ equal $b$. Let the proportion in which the inputs are at present combined equal $a/b$, and let $c = ka$, and $d = kb$ ($k < 1$). We know by definition that $\dfrac{c/k}{a} = 1$; $\dfrac{d/k}{b} = 1$.

Therefore        $c/d = a/b.$

From this it is easily seen that if $Fj$ and $Fi$ are assumed divisible in the sense here used, there must be constant returns to scale; since the rate at which $c$ can be substituted for $d$ is the same as that at which $a$ can be substituted for $b$, where $c$ and $d$ are equiproportionate amounts of $a$ and $b$.

---

1. The assumption is essential. If a hundred half-men are not equivalent to fifty whole men (Chamberlin, op. cit., p. 241), then there is no sense in which the fifty whole men can be said to have been divided. Arithmetic and economic meaning of units must correspond.

The above proposition is clearly one of definition only, and entails no "proof" of any sort. But definition is of primary and sole importance in this problem. Professor Chamberlin's difficulties seem to a large extent due to the fact that he does not realize that he must define a factor of production before he can say whether it is divisible or not.

His discussion of the practical possibilities of divisibility is a particularly good example of this fault. Professor Chamberlin considers that the continuous division of the general factor labor into more or less efficient laborers is an example of what we may mean "concretely" by divisibility.[2] Unfortunately he never defines his general factor labor, so that we do not know what it is that he is dividing. As far as I am aware no one has as yet supplied us with an adequate definition of the "general factor labor". It is in matters like these that mathematical applications, which Professor Chamberlin regards as irrelevant to the problem, are of some importance. If mathematical analysis is used, one cannot escape the necessity of providing a unique definition of the variables used. The meaning of a factor of production must be clear and unambiguous; hence for instance, the assumption nearly always made in mathematical analysis of production, etc., that labor is homogeneous. The formulation of divisibility which Professor Chamberlin attacks is indeed not a "result" of mathematical analysis, but a necessary premise on which such analysis must be based.

## II

Let us now examine some of Professor Chamberlin's contentions in a little more detail.

Professor Chamberlin maintains that economies of scale may in the "absence of indivisibilities" occur owing to (1) "increased specialization made possible in general by the fact that the aggregate of resources is larger, and (2) qualitatively different and technologically more efficient units of factors, particularly machinery, made possible by a wise selection from among the greater range of technical possibilities opened up by the greater resources."[3] But neither of these explanations appears to be satisfactory.

If there is perfect divisibility in the sense in which we have defined it, then it is possible to achieve the optimum degree of specialization for any level of output. This is quite clear *a priori*. For any given outlay of money it is possible to buy *exactly* the same

2. Op. cit., page 242
3. Op. cit., page 236

collection of inputs as with any greater outlay. Specialization has nothing whatsoever to do with size; since perfect divisibility is assumed, it is solely a matter of subdividing any single productive process into a large number of stages, and this by definition is possible, irrespective of the absolute amount of factors employed.

For instance, let us suppose that we have the choice of breeding men of the size and strength of ants, such that a hundred ant-men are equivalent to one representative standard-man in efficiency. Let us further suppose that there are ant-machines such that a hundred such machines are equivalent to one standard machine. Let the machines and men be the only factors (defined as above) employed. If there are, say, a hundred processes for which it is desirable to specialize one man and one machine, then it is clear that such specialization can be achieved either by one hundred men and one hundred machines at a large output, or by one man and one machine divided into one hundred ant-men and one hundred ant-machines at a smaller output. By definition a hundred ant-men in any one use are equi-efficient with one man in the same use, and similarly with machines. If, therefore, it is desired to produce a larger output after a time, each ant-man and-machine could be replaced by one man and one machine, or by fifty ant-men and fifty ant-machines, without one scrap of difference being made in the possible degree of specialization. Perfect divisibility leaves no room for increased specialization with increased size.

Similar criticism may be levied against the second explanation of economies of scale in the absence of indivisibilities. The only explanation of a "greater range of technical possibilities" with increasing scale of operation must lie in the indivisibility of these "technical possibilities." Indeed Professor Chamberlin's explanation is as good an example of indivisibilities as one could hope for. If a certain machine can only be used "at all economically" with a certain output, then it is clear that either it, or some other factor, is by our definition indivisible. Professor Chamberlin is able to talk of divisibility in this instance only by assuming that it is the generic factor "machinery" which is required to be divisible. But "machinery," like "labor," as a factor has no meaning. If there is to be divisibility, one *particular* machine must be divisible, and that in the sense in which we have defined it. Professor Chamberlin talks of "technologically more efficient units of factors." But what are the units in which he measures his heterogeneous collection?

Professor Chamberlin throughout his article is at great pains to

show that "size" is at least as important as "proportions" in the problem of returns, and that doubling *all* the factors of production, even though there are no indivisibilities and the proportions between the factors remain constant, need not necessarily lead to the doubling of output.  Apart from the criticism already made of his statement, the following further observations are relevant.

Increased (or decreased) specialization, which we have already noted to be excluded by the assumption of perfect divisibility, is also excluded by the assumption of constant proportions.  If we take Adam Smith's example of the pin makers and assume that there are ten workers using ten different tools before the increase in operations, and twenty identical workers with twenty tools, of which two are always identical, then it is clear that there is no possibility of increased specialization, if *all* the tools and workers are to be used.  The common sense proposition that the duplication of a given industrial process (together incidentally with a duplication of entrepreneurs or entrepreneurial function) will lead to a doubling of the output is unassailable.

If, by increasing the scale of operations, factors are brought into use which were not previously used, then this is not only incompatible with the statement that all factors are perfectly divisible (as we hope to have shown), but also incompatible with the statement that all factors are increased in the same proportion.   Any one factor, which is not used at a lower output cannot be used with a greater output, if for instance the latter is to be produced by doubling all the factors used at the smaller output.

Lastly, Professor Chamberlin seems to regard the fact that "the greater complexity of the producing unit as it grows in size . . . (leads) to increased difficulties of coördination and management"[4] as a proof that there is no most efficient proportion of factors independent of output, even if all factors are fully divisible and in variable supply.  This however is not the case.  Increased complexity is due to the fact that management and entrepreneurship are not increased in the same proportion as are the other factors, and to nothing else.  If management and entrepreneurship could be increased in the same proportion as the other factors there would be no increase in complexity.  If two identical entrepreneurs set up two identical plants, with an identical labor force, to produce an identical commodity $x$, then the two together will produce twice the amount of $x$ that could be produced by one alone.

4. Op. cit., page 249.

### III

In conclusion something more may perhaps be said concerning the nature of the subject in dispute in this note.

The question of what will be the effects of a perfect divisibility postulate on analysis, is mainly a question of definition. Controversies over definitions are mostly regarded as tedious and pedantic, yet there is much to be said for talking the same language. Also, at times one particular definition will enable analysis to be carried further, or to be conducted with greater ease, than will another. There is thus no great need to justify the topic under discussion.

Now the main difference between Professor Chamberlin's (implied) definition and the one here advocated, is that his pays lip-service to common usage (though not to common sense) with respect to such terms as labor and machinery, and to the concrete possibilities of "dividing" such generic groups; while our definition attempts to establish exact categories which can be used in analysis. It may be thought that his definition will enable him to analyze the problem of production on a plane less removed from reality and common sense than ours. This, however, is so only in one particular aspect of the problem.

Our definition confirms the commonsense view that the doubling of *all* factors of production must lead to a doubling in output, a view which Professor Chamberlin cannot reach since he has no exact statement of what he wants us to understand by "doubling all factors." On the other hand, Professor Chamberlin's view may be more realistic in the case of "halving" or generally dividing the factors of production. In our definition, probably very few factors are divisible in the real world, while Professor Chamberlin's concrete concept of dividing broad generic terms is more applicable. We hope however to have shown that this concrete view can hardly be used in rigorous analysis.

Lastly, a word should be said concerning the tautological nature of our conclusions. There is no harm in tautologies as such, indeed a great deal of modern economic analysis makes use of them — one need only look at modern welfare economics, the analysis of the firm, or expectations to confirm that view. A tautology is essentially a statement that cannot be disproved by an appeal to facts — it is an identity, a definition. Economics like any other discipline needs rigorous definitions. Care must, however, be taken not to define whole problems out of existence. This I think we are not guilty of, in using the definition here advocated.

Our definition of divisibility is essentially of the nature of a

standard of reference. If phenomena in the real world exactly fit that standard then by definition we shall expect constant returns to scale. Constant returns to scale is of course itself purely a matter of definition. I can see no valid objection to the use of such a standard of reference, nor do I believe that the problems Professor Chamberlin discussed are any other than problems of definition.

F. H. HAHN.

UNIVERSITY OF BIRMINGHAM, ENGLAND.

## REPLY

The arguments of both Mr. McLeod and Mr. Hahn provide striking examples of the familiar and widely-held thesis against which my earlier article was directed, and I believe that a complete and satisfactory answer to much of what they have said has already been given. But there are some misinterpretations of my own position to be cleared up, and some new lines of argument which afford the opportunity to strengthen and elaborate the case made earlier in several important directions.

One point of agreement should first be noted: we are agreed that the position I am attacking is held, not merely by the few economists I was able to mention in my earlier article, but very generally. Mr. McLeod speaks of it as "more in accord with common usage" than my own, and as "the customary approach"; Mr. Hahn twice links it with "common sense," describes it as "unassailable," and asserts that "the only definition feasible is one which identifies perfect divisibility with constant returns." Similar observations have been made to me by others, both orally and in written communications. Clearly it is not a straw man that is being attacked.

The issue of the real world *vs.* a theoretical "model" is raised by Mr. McLeod, who reproaches me, saying, "In effect he simply defines his terms to include only those types of divisibility that are practicable in the real world, and he discussed no other possibility."[1] To this I plead guilty, and with enthusiasm. I am sorry if I have, as he states later on, "missed an excellent opportunity of bridging the gap between theory and practice," for it was precisely my purpose to do something in this direction. My position was and is, in fine, that the theory of this subject is quite unnecessarily and absurdly remote from the real world, and that it should be brought into a closer relationship with it in order to be more useful in explaining this world instead of a world of ant-men. The answer, then, as to whether

1. Above, p. 129.

I am "speaking of the real world or of some particular theoretical model" is that I am speaking of (a) a "particular theoretical model" which is (b) modelled closely on the "real world." Economic theory needs the firm, because the firm is very important in the real economic world. And it needs both the falling and rising parts of the firm's cost curve for the same reason; explained in terms of the forces *actually* at work, instead of by arbitrary assumptions resorted to in order to put the economies and diseconomies back after they have been made to disappear by still other arbitrary assumptions.

We are told by both critics that the issue is simply and solely one of definition (a common view); and whether or not this is so will be discussed further on. But first it should be observed that fundamentally they hold the much more positive view that there is really no choice in the matter: the very definition of a (homogeneous) factor as a prerequisite to analysis allegedly carries with it the denial of economies and diseconomies of scale. Thus, according to Mr. Hahn, "the only definition (of divisibility) feasible is one which identifies perfect divisibility with constant returns,"[2] and my "difficulties" arise simply from sidestepping this crucial matter of defining a factor.

This is at bottom the "mathematical" argument: the "formulation of divisibility" which I attack is described as "a necessary premise (sic) on which such (mathematical) analysis must be based,"[3] and Mr. Hahn thus identifies himself with that school discussed in my article which seems to hold the strange opinion that the very use of mathematics commits one to some *particular* mathematical function. Mr. McLeod is less explicit, but in his last paragraph he speaks of my "refusal to discuss increments in the amount of a factor used by a given firm on the assumption that the units are of unvarying unit efficiency,"[4] and clearly holds also that to resolve the problem of variations in efficiency as between the "natural" units of a factor (individual laborers, for instance) by assuming a unit of "standard efficiency," is to be led inevitably to the "customary approach" in which economies and diseconomies of scale vanish.

This is certainly a very common view, and involves a mistake at such an elementary level that it perfectly illustrates the prevailing state of confusion in the subject. But first let me explain why a factor of production was not more explicitly defined in my earlier analysis.

There is not, I think, universal agreement as to how a factor of production should be defined, and I purposely refrained from adopting

2. Above, p. 131.
3. Above, p. 133.
4. Above, p. 130.

any *particular* definition in order to leave the argument more general, and hence acceptable, to those holding different views in the matter. Hence the discussion of fractional units was rather fully developed in order to cover numerous possibilities under different definitions of a factor and different conceptions of "dividing" it. The result was, as intended, that some of the discussion applies to one definition of a factor and of its division, and some to others. Its purpose, however, was always to look into the *actual* effect of divisibility upon efficiency under different circumstances and with different meanings given to the terms, and to show that there was *no* sense in which divisibility involved the collapse of the cost curve to a horizontal line — excepting, of course, when defined in a question begging way to include this result.

My own belief is that factors of production may be defined differently for different purposes, and that there is no single classification or definition of a "unit" which is universally and uniquely valid or most useful. For present purposes it is necessary only to accept the definition of my critics, as developed especially by Mr. Hahn, and to show that it does not, as represented, wipe out economies and diseconomies of scale.

Mr. Hahn tells us that "we must define a factor in a manner such that the marginal rate of substitution between its individual units is equal to unity for all levels of output."[5] Agreed. But it does not follow from this that efficiency is independent of output. To my knowledge no one has ever held that because individual oranges are interchangeable at the margin, the law of diminishing utility is denied. Similarly, the fact that Smith can be substituted for Jones in any aggregate of factors without effect upon the total product tells us nothing whatever about the shape of the production function. To interpret units of a factor of "standard efficiency" as meaning uniform efficiency *at all outputs* is as much nonsense as it would be to interpret it as meaning uniform efficiency *in all combinations with other factors*, i.e., as a denial of diminishing returns. One wonders why this latter has not been proposed.

Mr. Hahn would seem to agree, at one stage of his argument, that efficiency is not necessarily independent of output, when, in discussing fractional units of a single factor he points out explicitly that the fact that the marginal rate of substitution between $1/k$ the amount $ka$ and $a$ is unity "is, of course, perfectly compatible with either increasing, constant or diminishing returns to successive

5. Above, p. 132.

amounts $ka$ of the category $Fi$."[6]  But with the addition of a second factor and the manipulation of a few more symbols, he concludes that "it is easily seen that . . . there must be constant returns to scale."  In truth, all that the mathematics shows is that the ratio of (say) 1/3 of factor $a$ to 1/3 of factor $b$ is the same as that of $a$ to $b$, which could hardly be disputed.  It is also perfectly compatible with either increasing, constant or diminishing returns to successive amounts of the two factors together (either in the fixed proportion $a/b$ or, possibly, with varying proportions).  Indeed, the cancellation from which emerged the equation $c/d = a/b$ even removes the earlier requirement that there be no loss of efficiency through dividing the factors.  Such loss is consistent with the final result provided only that it be the same proportionally for both factors.

Mr. Hahn's argument in its various phases rests heavily upon the concept of fractional units, and it should be called to mind therefore that this may be a minor part of the whole problem.  The argument is made in my earlier article[7] that without considering fractional units at all the curve is defined for practical purposes by the discontinuous series of dots at the unit breaking points.  It may now be added that variations in the efficiency of the "natural units" may in itself result in virtual continuity with respect to "efficiency units."  A simple illustration is found in an analogy with dividing a sum of money.  With only ten pieces of money: one half-dollar, one quarter, one dime, two nickels and five pennies, any whole sum between one cent and a dollar may be had without resorting to fractions of coins.  Similarly, if the efficiency of Jones is taken as unity, and we find that Smith is equal to 1.13 efficiency units, White to .84, Brown to 1.07, etc., etc., the finest gradations on the scale of efficiency units may be possible without ever dividing a natural unit at all.  Such considerations are especially important where natural units *cannot* be divided in any meaningful sense, and therefore where it is futile to discuss what the efficiency of a fractional unit would be.[8]  In any event, the assumption of a standard efficiency unit as a measuring rod for quantitative analysis, when considered together with the actual heterogeneity of natural units, far from collapsing the cost curve to a horizontal line, will instead merely make it more continuous than before, and therefore reduce still further the already limited importance of the issue as to what dividing a unit does to its efficiency.

Perhaps considerations of this sort would spare us the ant-men,

6. Above, p. 132.
7. Loc. cit., p. 240.
8. Cf. my earlier article, pp. 244–45.

although I think we ought to be spared them on general grounds as well. My own earlier mention of division by meat cleavers and the like was expressly in order to turn attention away from such silly examples to economic realities. Yet Mr. Hahn appears to be quite serious, and from his excursions into ant-economics emerges with the conclusion that "perfect divisibility leaves no room for increased specialization with increased size."[9] Let me then be equally serious in pointing out what is wrong. There is no simple and single interpretation to be given to the "division" of a factor; and if Blackstone the magician could establish the necessary link between the human and the ant world by causing Jones to disappear in a cloud of smoke from which would emerge 100 ant-men, this result would no more illustrate "perfect" divisibility than would the use of the meat cleaver or the employment of part-time workers. In all these cases, including the ant-men, it is still necessary to inquire: *what does the division do to efficiency?* I have never met an ant-man, and therefore cannot give the answer, but I am very skeptical as to whether even 100 of them could do the work of Jones. The point is that such weird concepts do not illustrate "perfect divisibility"; they only serve as a device for begging the question by *assuming* unchanged efficiency as a part of the definition of an "ant-man."

Mr. Hahn twice refers to the "commonsense" view that doubling all factors must double output, and adapts Adam Smith's example of pin manufacture to demonstrate that with twice as many workers and tools "there is no possibility of increased specialization."[1] Yet what could be more sensible and more in accord with the most elementary economics than this: that with more workers and tools the division of labor could be carried farther! If an office staff expands from five to ten secretaries, always of equal efficiency, each complete with typewriter and desk, is there not typically a further division of labor between them, each acquiring an added proficiency in some particular part of the total work to be done? Again, ten carpenters, each with a standard kit of tools, could certainly organize the manufacture of chairs amongst themselves by division of labor so that the output would be more than ten times that of a single carpenter and kit. Of course in real life there are typically also qualitative changes in the factors to be taken into account with any substantial change in their amounts, but what "common sense" is there in supposing that a restrictive assumption which would bar such changes would thereby remove *all* possibilities of specialization?

With reference to coordination and management, Mr. Hahn gives

9. Above, p. 134.
1. Above, p. 135.

no criticism of my argument, but merely reasserts that "increased complexity is due to the fact that management and entrepreneurship are not increased in the same proportion as are the other factors and to nothing else."[2]   Evidently he believes that an army is no more complex than a squad, no higher controls being needed to hold it together if only every squad is headed by its corporal.   But if one grants that more "coordination" relative to other factors *is* needed with a larger control unit, the conclusion reached in my earlier analysis, that failure to provide it by holding proportions constant only makes matters worse, is unescapable.   At the close of his discussion of this problem it becomes clear that he is confusing the expansion of an *industry* by the addition of identical firms, with the expansion of the firm itself.

Both of my critics defend tautologies, and Mr. McLeod even finds that my own argument "may also be dismissed as tautological.[3] As I seek to discover why, it appears to be because I have rejected tautological definitions, and thereby "exclude the influence of indivisibility in the sense used by others."   Now if rejecting tautological definitions really is in itself tautological, I am bound to say that there appears to be no escape.   But Mr. McLeod has confused two things:   (a) the rejection of any definition of divisibility which *requires* that the efficiency of a factor be not affected, which I have done, and (b) the rejection of any definition which *permits* that efficiency be not affected (thereby assuring by definition that it *will* be affected), which I have not done.   My inquiry was directed to the question:   "What, if anything, *does* divisibility do to efficiency?"[4] and the appeal is constantly to the facts of the real world.   Among the conclusions is that it *may* do nothing at all, *the cost curve then remaining U-shaped for other reasons.*[5]   There are no *assumptions* as to efficiency included in any of the several definitions discussed, and hence there is no begging of the question at issue.

Mr. Hahn comments that "there is no harm in tautologies, indeed a great deal of modern economic analysis makes use of them — one need only look at modern welfare economics, the analysis of the firm or expectations. . . "[6]   This, I fear, is only too true, and goes far to account for the emptiness and futility of a great deal that currently passes for "economic theory."   However, it is not "rigorous definition" which is at fault, but, in Mr. Hahn's own words, definitions

2. Above, p. 135.
3. Above, p. 129.
4. Loc. cit., p. 239; see also top of p. 249.
5. Loc. cit., p. 240
6. Above, p. 136.

which "define whole problems out of existence." This latter might serve as a rough description of "tautology" in our present context; it is certainly not to be identified, as Mr. Hahn asserts, merely with "definition." In the matter at hand, the economist's responsibility is to explain the phenomenon of economies of scale; and in his standard reply he solemnly asserts that his study of the problem has led him to the conclusion that there will be such economies only if there are such economies. This, as I understand it, is tautological, and, although it may not be "wrong," it is singularly lacking in nourishment.

I am certain that some will share the view repeatedly expressed by Mr. Hahn that the "whole problem" is one merely of definition; and as an aid to the reader in deciding whether it is or not, I should like to put before him a story attributed to Abraham Lincoln. When visited by a delegation of clergymen in 1862 urging him to proclaim the slaves free, he replied by asking them how many legs a sheep would have if you called its tail a leg. "Five," they replied at once. "You are mistaken," said Lincoln, "for calling a tail a leg doesn't make it so." If some visiting Martian should ask how many legs our animal the sheep had, it would appear to me quite wrong to reply: "It is simply and solely a matter of definition; for instance, if the ears are included . . . but on the other hand, if only the tail. . . etc." Unless it can be shown that "perfect divisibility" (*in some non-question-begging sense*) removes economies and diseconomies of scale, there would seem to be nothing but error in attributing such economies and diseconomies to "imperfect divisibility." To call them thus doesn't make it so; and, as I hope to have shown earlier, it has actually led to wide ramifications of further erroneous analysis.[7]

<div align="right">Edward H. Chamberlin</div>

Harvard University

7. I should like to add a word of comment on Mr. Allen's brief discussion of the problem in relation to the formulations of Hicks and Samuelson. (Above, p. 116 and note.) It is perfectly true, of course, that (a) if all factors are not included (Samuelson), or (b) if one factor is held fixed (Hicks), an explanation is afforded of why "doubling the use of all factors does not double the output." But to admit only these possibilities seems to me dangerous and misleading, for it implies that *if* all factors were (a) included and (b) variable, the function would be linear and homogeneous, which I maintain is quite without foundation. If such expedients as leaving out factors or holding some of them fixed are necessary in order to "make the curve turn up," I predict that the linear homogeneous production function, which we seem to agree should be abandoned, will have a long and healthy existence; for such a position constitutes a tacit admission that with all factors present and continuously variable (which seems to be the most "general" assumption) the function *is* homogeneous. It evidently involves also a completely different explanation of the *falling* part of the curve from that given in my analysis.

# RECENT PUBLICATIONS

AFTALION, ALBERT. Monnaie et Economie Dirigée. Paris, France: Librairie du Recueil Sirey. 1948. pp. 414.

BLACK, JOHN D., AND KIEFER, MAXINE E. Future Food and Agriculture Policy. New York: McGraw-Hill Book Company. 1948. pp. 348. $3.50.

BODDY, FRANCIS M., CHILDES, FRANK E., ET AL. Applied Economic Analysis. New York: Pitman Publishing Corporation. 1948. pp. 573. $3.75.

BONN, M. J. Wandering Scholar. New York: The John Day Company, Inc. 1948. pp. 403. $5.00.

BUEHLER, ALFRED G. Public Finance. New York: McGraw-Hill Book Company. 1948. pp. 740. $5.00.

CAREY, JANE PERRY CLARK. The Role of Uprooted People in European Recovery. Washington, D. C.: National Planning Association. 1948. pp. 85. $1.00.

CHASE, STUART. The Proper Study of Mankind: An Inquiry into the Science of Human Relations. New York: Harper & Brothers. 1948. pp. 311. $3.00.

DAUGHERTY, CARROLL R. Labor Problems in American Industry. Cambridge, Mass.: The Riverside Press. 1948. pp. 1066.

DIETERLEN, PIERRE, AND RIST, CHARLES. The Monetary Problem of France. New York: King's Crown Press. 1948. pp. 98. $2.50.

FOOD AND AGRICULTURE ORGANIZATION OF THE U. N. The State of Food and Agriculture — 1948. New York: Columbia University Press. 1948. pp. 216. $2.00.

HARRIS, SEYMOUR E. The European Recovery Program. Cambridge, Mass.: Harvard University Press. 1948. pp. 309. $4.50.

HARRIS, SEYMOUR E., ED., Foreign Economic Policy for the United States. Cambridge, Mass.: Harvard University Press. 1948. pp. 490. $6.00.

KAPLAN, A. D. H. Small Business: Its Place and Problems. New York: McGraw Hill. 1948. pp. 281. $3.25. C. E. D. Research Study.

KERR, CLARK AND RANDALL, ROGER. Causes of Industrial Peace under Collective Bargaining: A Study of Crown Zellerbach and the Pacific Coast Pulp and Paper Industry. Berkeley, California: Institute of Industrial Relations, University of California. 1948. pp. 78. $1.00.

KILLOUGH, HUGH B. AND KILLOUGH, LUCY W. Economics of International Trade. New York. McGraw-Hill Book Company. 1948. pp. 463. $5.00.

LESTER, RICHARD A. Company Wage Policies: A Survey of Patterns and Experience. Princeton, New Jersey. Industrial Relations Section, Princeton University. 1948. pp. 45. $1.50.

LEWIS, CLEONA. The United States and Foreign Investment Problems. Washington, D. C. The Brookings Institution. 1948. pp. 359. $4.00.

LINDHOLM, RICHARD W. Introduction to Fiscal Policy. New York: Pitman Publishing Corporation. 1948. pp. 235. $2.75.

——. Public Finance of Air Transportation. Columbus, Ohio: Ohio State University. 1948. pp. 178.

LINTNER, JOHN. Mutual Savings Banks in the Savings and Mortgage Markets Cambridge, Mass.: Harvard Business School. 1948. pp. 559. $6.50.

LOUCKS, WILLIAM N. AND HOOT, J. WELDON. Comparative Economic Systems. New York: Harper & Brothers. 1948. pp. 836.

MILLER, GLENN W. American Labor and the Government. New York: Prentice-Hall, Inc. 1948. pp. 638. $7.35.

MORGAN, THEODORE. Hawaii. Cambridge, Mass.: Harvard University Press. 1948. 260 pp. $4.00.

MYINT, HLA. Theories of Welfare Economics. Cambridge, Mass.: Harvard University Press. 1948. pp. 240. $4.00.

STATE OF NEW YORK, DEPARTMENT OF LABOR. Discharge for Cause. New York: Division of Research and Statistics. 1948. pp. 88. 70 cents.

PATTERSON, S. HOWARD AND SCHOLZ, KARL W. H. Economic Problems of Modern Life. New York: McGraw-Hill Book Company. 1948. pp. 782. $4.75.

PETTENGILL, ROBERT B. Price Economics. New York: The Ronald Press Company. 1948. pp. 483. $4.50.

PILLIET, GEORGES. Inventaire Economique de la France. Paris, France: Les Ordres de Chevalerie. 1948. pp. 417. $3.00.

POLLAK, OTTO. Social Adjustment in Old Age. New York: Social Science Research Council. 1948. pp. 199. $1.75.

SLICHTER, SUMNER H. The American Economy. New York: Alfred A. Knopf, Inc. 1948. pp. 214. $2.75.

SOULE, GEORGE. Introduction to Economic Science. New York: The Viking Press. 1948. pp. 154. $2.50.

UNITED NATIONS. Balance of Payments, 1939-1945. Geneva, Switzerland: United Nations (U. S. agents: Columbia University Press). 1948. pp. 207. $2.50.

WILKS, S. S. Elementary Statistical Analysis. Princeton, New Jersey: Princeton University Press. 1948. pp. 294. $2.50.

WITHERS, WILLIAM. Public Finance. New York: American Book Company. 1948. pp. 489. $4.25.

WOLMAN, LEO. Industry-Wide Bargaining. Irvington-on-Hudson, New York: The Foundation for Economic Education. 1948. pp. 63. 50 cents.

WOOD, RICHARD H., AND PEARCE, JOHN J., JR. The Joint Safety Program of The Forstmann Woolen Company and Local 656 Textile Workers Union of America, C.I.O. New Brunswick, New Jersey: Rutgers University. 1948. pp. 47. 40 cents. [Case Studies of Cooperation between Labor and Management: No. 1.]

# American Economic Review

## CONTENTS

# THE

# QUARTERLY JOURNAL

# OF ECONOMICS

## VILFREDO PARETO
## (1848-1923)

SUMMARY

Introduction, 147. — I. The Man, 149. — II. The Theorist, 153. — III. The Sociologist, 167.

In a volume devoted to Pareto's life and work,[1] Professor Bousquet relates that the obituary article devoted to Pareto in the socialist daily, *Avanti*, described him as the "bourgeois Karl Marx." I do not know that a man can be rightly called "bourgeois" who never missed an opportunity to pour contempt on *la bourgeoisie ignorante et lâche*. But for the rest, the analogy conveys very well the impression that Pareto had made upon his countrymen: they had in fact raised him to an eminence that was unique among the economists and sociologists of his time. No other country erected a similar pedestal for his statue, and in the Anglo-American world both the man and the thinker have remained strangers to this day. There was, indeed, a short Pareto vogue in this country that followed upon the translation of his sociological treatise.[2] But it died out soon in

1. G. H. Bousquet, "Vilfredo Pareto, sa vie et son oeuvre," (in the *Collection d'études, de documents. et de témoignages pour servir a l'histoire de notre temps*, Paris, Payot, 1928). Except for the mathematical parts of Pareto's work, this book, written in a vein of generous enthusiasm by a man who is an economist and sociologist in his own right and as far as possible removed from the state of mind of the disciple or biographer who basks in reflected glory, is herewith strongly recommended. Bousquet also wrote a *Précis de sociologie d'après Pareto*, introductions to the latter's *Systèmes Socialistes* and *Manuel d'économie politique*, and also a short English appraisal entitled, *The Work of Vilfredo Pareto*, 1928, besides reserving for him a place of honor in his *Essai sur l'évolution de la pensée économique*. Of other memorial appraisals it will suffice to mention what may be called the official one, Professor Alfonso de Pietri-Tonelli's address to the economic section of the Italian Association for the Advancement of Science, published in the *Rivista di Politica Economia*, November and December, 1934, and January, 1935, and Professor Luigi Amoroso's article in *Econometrica*, January, 1938.

2. At Harvard, this vogue was represented by the eminent physiologist, the late Professor L. J. Henderson. See his *Pareto's General Sociology*, 1935. Some Harvard men will still remember his informal Pareto "seminar" that practically consisted in a series of monologues by the Professor. Sympathetic understanding and a profound sense of the unconventional greatness of Pareto's thought there struggled valiantly with inevitable professional handicaps.

an uncongenial atmosphere. Moreover, so far as the small circle of pure theorists is concerned, Pareto came to exert considerable influence on Anglo-American economics in the 1920's and 1930's, that is, after the publication of Professor Bowley's *Groundwork*. But both in England and the United States, Marshallian and post-Marshallian economics offered enough in the line in which Pareto excelled to prevent him from gaining much ground of his own even before other tendencies took away whatever he had gained.

This might seem surprising owing to the fact that several important developments in theoretical economics are now seen to stem from him. But it is not difficult to explain. Pareto was the product of a sector of the Franco-Italian civilization that is far removed from English and American currents of thought. Even within that sector his towering figure stood almost alone. Pareto cannot be pigeonholed. He paid court to no "ism." No creed or party can claim him as its own, although many creeds and parties appropriated fragments of the vast intellectual realm over which he held sway. He seems to have taken pleasure in running counter to ruling humors and slogans. Votaries of extreme laissez-faire may cull plenty of passages from his writings in support of their views. Yet there was nothing he despised so thoroughly as the "pluto-democracy" or "plutocratic demagogy" of liberalism. Socialists are under obligation to him for rendering, as we shall see, a very important service to socialist doctrine, and also for his protests against the anti-socialist measures that the Italian government took in 1898. Yet he was not only an anti-socialist but one of that type whose criticism derives sting from contempt. French Catholics might thank him for his attacks upon the persecution of the French clergy that was so unedifying a sequel to the Dreyfus affair. Yet he attacked the "laicist" policies of the Combes ministry because he was a gentleman, and not because he believed either in the mission of the Catholic church or in her teaching.

A gentleman of such independence and pugnacity who is in the habit of dealing vigorous blows right in the midst of arguments that might in themselves be agreeable to some party or another has little chance of being popular. By now he is a figure of the past. But even at the epoch of his prime the political and social slogans with which we are all familiar controlled official phraseology, the press, party programs, and popular literature including its economic sector. The wrapping in which he presented his strictly scientific results were then not much more popular than they would be now. One has only to imbue oneself with the spirit that pervades an American textbook

and then to open Pareto's *Manuel* in order to realize what I mean: the naïve lover of modern social creeds and slogans must feel himself driven with clubs from Pareto's threshold; he reads what he is firmly resolved never to admit to be true and he reads it together with a disconcerting wealth of practical examples. Therefore it seems that the problem is not to explain why Pareto did not exert influence more widely; the problem is rather to explain how he came to exert as much as he did.

Could we confine ourselves to Pareto's contributions to pure theory, there would be little need for glancing at the man and his social background and location. But into everything that was not a theorem in the pure logic of economics the whole man and all the forces that conditioned him entered so unmistakably that it is more necessary than it usually is in an appraisal of scientific performance to convey an idea of that man and of those forces. I shall make an attempt to do so first (I). Then I shall briefly survey Pareto's work in pure theory (II). And I shall end up with a glance at his conception of society that has found so inadequate an expression in his *General Sociology* (III).[3]

## I. THE MAN

Pareto's father, the Genoese Marchese Raffaele Pareto, seems to have been a typical product of the Italian Risorgimento of the first half of the nineteenth century, an ardent adherent of Mazzini — perhaps more from national than from social reasons — an uncompromising enemy of all the governments that barred Italy's way towards national unity and a revolutionary in this if in no other sense. Accordingly, he exiled himself to Paris where Vilfredo, the

3. There is a bibliography that cannot be very far from being complete by Messrs. Rocca and Spinedi in the *Giornale degli Economisti*, 1924, but only the following items need be mentioned here: "Considerazioni sui principi fondamentali dell' economia politica pura," *Giornale degli Economisti*, 1892–3; *Cours d'économie politique professé a l'université de Lausanne*, 1896–7; *Résumé du cours donné a l'Ecole des Hautes Etudes Sociales de Paris*, 1901–2; *Les systèmes socialistes*, 1902 (reprinted 1926); *Manuale di economia politica*, 1906 (reprinted 1919); *Manuel d'économie politique*, 1909 (reprinted 1927) (a translation of the preceding item which must however be listed separately because of the mathematical appendix that was completely redone); *Trattato di sociologia generale* (1916) French translation, 1919, English translation, under the title *Mind and Society*, 1935; "L'économie mathématique" in the French *Encyclopédie des sciences mathématiques*, 1911 (the corresponding article in the original German edition of the Mathematical Encyclopedia is of negligible importance). There are several other books besides innumerable articles but they do not, so far as I know them (Pareto published many articles in the daily press, most of which I do not know), contain anything of a scientific nature that is not contained in one or more of the publications mentioned.

subject of this memoir, was born of a French mother: if General Galliéni once described himself as *"Francese ma anche Italiano,"* Vilfredo Pareto might have described himself as *"Italiano ma anche Francese."* He was taken to Italy in 1858 and there went through the usual course of studies that issued in a Doctor's degree in engineering in 1869. He immediately embarked upon engineering and industrial management as a profession and after various other appointments rose to be manager general — we should say "president" — of the Italian Iron Works. It was only in 1893 that he was appointed successor of Walras in the University of Lausanne, although he may be considered as a full-time economist a few years before that. Thus, the span during which he was primarily engaged in economic research extends from about 1892 to about 1912 — practically all his later work is sociological in nature. He resigned his chair in 1906 and then retired to his home, a country place on the lake of Geneva, to grow in the course of a vigorous and fertile old age into the "lone thinker of Céligny."

Substantially, this suffices for our purpose: we have to underline a few of these facts rather than to add others. First, theorists will note that owing to his training as an engineer — and he seems to have cultivated theoretical aspects — he acquired at an early age command of mathematics on a professional level.[4] Second, it is worth-while to notice that, to a degree quite unusual with scientific economists, Pareto was thoroughly familiar with industrial practice — familiar in a sense which is quite different from the kind of familiarity that may be acquired by the means available to the academic economist, the public servant, the politician. But, third, it was his passionate interest in the current issues of economic and general policy, presently to be commented on in another connection, which made him something of an economist long before he started his own creative work. Francesco Ferrara was then at the height of his fame and influence, and the frosts had not yet fallen upon a theoretical structure glorified by uncritical liberalism. His writings, especially his famous introductions (*prefazioni*) to the classics published in the *Biblioteca dell' economista,* served Pareto as well as, or better than, any of the university courses could have done that were available in his student days. His way to Walras, however, was chalked out later on by Maffeo Pantaleoni.

4. I feel unable to say precisely how much this amounted to. Pareto had to be told by Volterra that an expression of the form $Xdx + Ydy$ has always an infinity of integrating factors whereas with more than two variables no such factor *need* exist. (*Manuel,* p. 546n.). I wonder whether a real "professional" could have overlooked this.

Nothing of the above will account completely for Pareto's vision of society and politics, or even for his attitudes to the practical problems of his age and country. Nor do I believe for a moment that the deep pool of personality can ever be drained so as to show what is at the bottom of it. But there is the patrician background which, I am sure all who knew him will agree, meant much more in his case than it usually does. In particular it prevented him from becoming a brother in spirit to the men — and a fully accepted member of the various groups — with whom life threw him. It also prevented him from establishing *emotional* relations with the creations of the bourgeois mind, such as the twins that are called democracy and capitalism. Acting upon this background, his financial independence — a bare independence at first, something like affluence later on[5] — helped to isolate him still further by offering the possibility to do so.

Still acting upon this background, his classical scholarship worked the same way. I do not mean that part of it which he shared with every educated person of his time, but that part which he conquered himself through incessant study of the Greek and Roman classics during his sleepless nights. The ancient world is a museum, not a laboratory of applied science, and he who trusts too much to the wisdom to be gathered there is bound to stray from every group of men that was in existence either in 1890 or in 1920. Isolation was made complete by the result of his participation in the debates on the policies and politics of his country — so complete that he had decided to emigrate to Switzerland even before he received the call to Lausanne. And isolation had its effects — soothed only late in life by a second marriage that brought domestic peace (see the dedication of the *Trattato*) — upon a fiery temperament that was not really made to stand it.

But why should he have left his country in wrathful exasperation — the country that he loved from the bottom of his heart and whose national rebirth he had not only yearned for but witnessed? The detached observer is all the more likely to ask this question because it will seem to him that the new national kingdom did not do so badly in the thirty years that preceded Pareto's emigration. Besides progressing economically at a considerable rate and growing out of financial disorders — *pace* our Keynesians — it took its first steps in social legislation and established itself successfully as one of what then were called the great powers. Looking at things in this

5. This relative affluence was due to an inheritance, not to his previous activity as a business executive.

light, our observer will develop a good deal of respect for a régime such as that of Agostino Depretis. And, considering the difficulties incident to the beginnings of the new national state, he will make allowances for the less exhilarating parts of the picture. But Pareto made no such allowances. He saw nothing but incompetence and corruption. He fought with impartial ferocity the governments that succeeded one another, and it was then that he became known as an ultraliberal — in the nineteenth-century sense of uncompromising advocate of laissez-faire — and that he helped to create, among the German New-Dealers of that period, the impression that marginal utility was just a wicked trick with which to thwart reformers.[6]   Possibly this is all there is to say about Pareto's attitude in matters of economic policy and the strong traces it left upon his scientific writing before 1900.  But even then there was something in that ultraliberalism of his that points in a direction exactly opposite to the creeds and slogans of official liberalism.  He certainly was *anti-étatiste*, but for political reasons rather than for purely economic ones: unlike the English classics, he did not fight against government activity *per se* but against the governments of parliamentary democracy, of the very parliamentary democracy that commanded the fervent allegiance of the English classics.  Viewed from this angle, his type of laissez-faire acquires a connotation that is entirely at variance with the laissez-faire of the English type.  And once we realize this, the rest is easy to understand.

Toward the end of the nineteenth and during the first two decades of the twentieth century, an increasing number of Frenchmen and Italians began to voice dissatisfaction that varied from mere disappointment to violent disgust at the manner in which the *cotillon* of parliamentary democracy functioned and at the results it produced in France and Italy.  Such sentiments were shared by men so different as E. Faguet and G. Sorel, and they were not confined to any one party.  This is not the place to analyze let alone to pass judgment upon them.  All that matters for us is their existence and the fact that the later Pareto stands out from this current of thought only because he himself stood out from his contemporaries and because he wrote a sociology that was — along with those of Sorel and Mosca — to rationalize it.

Englishmen and Americans, oblivious of the particular and

6. German critics received his *Cours* accordingly. In fact, the book contains very little that may be used for a different interpretation. It does, however, contain the remark that the virtues therein predicated of pure competition have no bearing upon the actual economic process since pure competition does not actually prevail.

historically unique circumstances that have developed in their minds an equally particular and unique attitude toward parliamentary democracy, have wondered about the possible meaning of Pareto's attitude towards Fascism. But this attitude is not problematical in the least. No theory is necessary in order to explain it. The events of 1914–1922 had called him back to the arena of political debate. The masterly analyses he published on the origins of the First World War, on the miscarriage of Versailles and on the futility of the League are among his strongest performances although they found no answering echo outside of Italy. But above all he witnessed with something like horror the social disorganization in Italy which it is necessary to have seen in order to believe. Attributing all the troubles of those years to the weakness of the political system of a decadent bourgeoisie, the student of Roman history may have thought of the formula by which, in republican Rome, the senate, in order to deal with an emergency, used to direct the consuls to appoint an officer of practically unlimited though temporary power, the dictator: *videant consules ne quid detrimenti res publica capiat.* But there was no such provision in the Italian constitution, and it would not have done any good if there had been one. So the dictator had to appoint himself. Beyond this and beyond approval of the success with which Mussolini restored order, Pareto never went. Mussolini honored himself by conferring senatorial rank on the man who kept on preaching moderation and who stood throughout for the freedom of the press and of academic teaching.[7] But to his last day Pareto refused to embrace this "ism" as he had refused to embrace any other. There is no point whatever in judging his action — or, indeed, any action or sentiment of his — from the standpoint of Anglo-American tradition.

Everything else is at the bottom of the pool.

## II. The Theorist

Any appraisal of Pareto's contribution to economics must first of all give due credit to a feat of leadership. He never taught in Italy. The faculty of Law in the University of Lausanne did not make very favorable headquarters for a campaign of intellectual conquest. The country house in Céligny looked like a *buen retiro*. Yet he did what Walras had not been able to do: he formed a school in the full sense of the word. An inner circle of eminent economists, a wider circle of followers of less eminence, and beyond this a broad fringe of more or less definite adherents emerged soon after 1900.

7. See on this, Bousquet, *op. cit.*, pp. 182–194.

They co-operated in positive work. They cultivated personal contact. They stood for one another in controversy. They recognized One Master and One Doctrine.

This school was specifically Italian. As has been pointed out already, there were but few foreign adherents though individual pieces of Paretian teaching eventually gained acceptance both in England and in the United States. Nor did the Pareto school ever dominate Italian economics. No·school ever does dominate its own country. Impressions to the contrary, e.g., the impression that the Ricardo school ever dominated English economics, are due to nothing but unrealistic historiography. Several other Italian leaders, like Einaudi, held their own ground entirely, and others, like Del Vecchio, while recognizing Pareto's eminence and adopting this or that of his doctrines, thought and wrote pretty much as they would have done had Pareto never lived. Still, there remains the fact that a school did emerge on the basis of a theoretical structure that was inaccessible not only to the general reader but, in some of its most original parts, also to students of economics, students moreover who had never heard or seen the master.

But once we have duly recognized and thereupon discarded this feat of leadership, we see a theorist who continued the work of Walras. Nobody, of course, ever denied this, not even the most ardent disciple and, least of all, Pareto himself. Difference of opinion on this point is inevitably confined to the extent to which he surpassed the great pioneer and to the relative mental statures of the two men. There are several reasons why disciples will never agree on this either with outsiders or among themselves. One of these reasons must be noticed at once. Walras presented his immortal theory in the garb of a political philosophy that is extra-scientific in nature and, moreover, not to everyone's taste. I am afraid that there is no better way of conveying what that philosophy was than to call it the philosophy of petty-bourgeois radicalism. He felt called upon to preach an *idéal social* that hails from the semi-socialist French writers of the first half of the nineteenth century or, as we may say with equal justice, from utilitarianism. He looked upon the nationalization of land as an essential item in his teaching, and he was a monetary reformer whose plans have a strikingly modern ring. All this was gall and wormwood to Pareto. It was just metaphysical speculation and metaphysical speculation of a very unsympathetic kind. Their common ground was confined to pure theory and specifically to Walras' equations of equilibrium. But in every other respect they were as different as two men can be, and even their

companionship in arms in the fight for mathematical economics and Pareto's obligation to Walras in the matter of the Lausanne professorship did not prevent their deep-seated mutual dislike from asserting itself or even from spilling over in conversation with third persons. While their pure theories are cast in the same mould, their systems of thought taken as wholes and their visions of the social process are not. And all those economists who are not disposed to neglect a man's philosophy and practical recommendations completely, that is to say the majority of the profession, will, for this reason alone, consider the Paretian structure to be something completely different from that of Walras.

In any case — we are neglecting sociology for the moment — it was, with one exception, in pure theory alone that he made scientific history. Let us note this exception first. In the *Cours* and also in a separate memoir of 1896 Pareto published a highly original pioneer achievement in econometrics that first established his international reputation and, under the title of "Pareto's Law," created what may be fairly called a whole literature devoted to its critical discussion. Call $N$ the number of income receivers who receive incomes higher than $x$, and $A$ and $m$ two constants; then Pareto's "Law" asserts that

$$\log N = \log A + m \log x$$

Chapter 7 of the *Manuel* contains Pareto's most mature interpretation of this generalization. We must confine ourselves here to noticing the two classes of problems which it raises. There is, first, the question of fit. Numerous investigations have been made some of which were held by their authors either to refute the Law completely or else to establish the superiority of other methods of describing the inequality of incomes. The reader will observe that the central issue turns on the approximate constancy of the $m$. However, by and large, the "Law" has stood fire rather well as the fact suffices to prove that it is sometimes used by competent statisticians even now. But there is, second, the question of interpretation. Granted that up to quite recent times the distribution of incomes according to brackets has been remarkably stable what are we to infer from this? This problem has never been attacked successfully. Most participants in the discussion, Pigou among them, have confined themselves to criticizing Pareto's own interpretation — which, to say the least, was in fact open to objection at first — and, like so many of our controversies, this one has petered out without yielding a definite result. Few if any economists seem to have realized the

possibilities that such invariants hold out for the future of our science.[8]  Viewed from this standpoint, Pareto's "Law" is path-breaking in the literal sense even though in the end nothing whatever is left of its particular form.

I take this opportunity to dispose of another matter.  In the *Manuel*, Pareto dealt with his "Law" of Income Distribution in the chapter on Population.  So far as the topics are concerned that are usually dealt with under this heading, this chapter does not contain much that would call for notice.  But it contains a number of other things which, like the "Law" are not usually included in the theory of population, and it is these items which liven up this chapter and impart to it its freshness and originality.  Pareto's theory of the *circulation of the élite* is an example (see below, section III).  Most of them are sociological rather than economic in nature, and some of them bring out sharply, indeed almost naïvely, certain prejudices that sat so incongruously upon the great analyst of human prejudice.[9]

In the field of pure theory, properly so called, Pareto's thought developed slowly and in fact retained certain pre-Paretian features to the end.  In addition to the early influences of Ferrara and of the English and French economists of the "classic period," he had Walras' equations of static equilibrium to start from — after having realized, not without considerable initial resistance, that they were in fact the keys to everything else.  He was further stimulated by all the suggestions that no competent theorist could have helped receiving in the decade from 1885 to 1895[1],  Finally, he was acutely aware of the technical shortcomings and other limitations of his immediate predecessors.  Thus his own theoretical work was cut out for him —

8. In particular, nobody seems to have realized that the hunt for, and the interpretation of, invariants of this type might lay the foundations of an entirely novel type of theory.

9. For instance, whatever we may think of his explanation of the phenomenon of feminism, we can hardly suppress a smile when we read the first sentence of this discussion (p. 400) which begins with the words, "Le féminisme est une maladie. . ." a turn of phrase that does not indicate much objectivity or detachment.  Both on Pareto's theory of population in the strict sense and on those sociological additions, I refer the reader to Professor J. J. Spengler's "Pareto on Population," this *Journal*, August and November, 1944.

1. To some of these suggestions Pareto reacted in a negative, if not, indeed, in a hostile manner.  He never appreciated Marshall fully — mainly because he objected on principle to partial analysis — and he seems never to have seen all there was behind the primitive technique of the Austrians.  But he did appreciate Edgeworth and, many objections notwithstanding, Wicksteed.  Much more than is commonly known he appreciated Irving Fisher, not only the *Mathematical Investigations* but also, later on, the *Nature of Capital and Income* and the *Rate of Interest*.  It was a revelation to me to hear him bestow high praise on *Capital and Income*.

most of it, indeed, by Walras himself.[2]  But his earlier work, such as his "Considerazioni sui principi fondamentali dell' economia politica pura" (*Giornale degli Economisti*, 1892–3), never went beyond the range of the Walrasian signposts.  This is also true, and emphatically so, of his *Cours*.  Some economists who respected Pareto but were not strictly Paretians have paid him the dubious compliment of calling the *Cours* his masterpiece.  It is, indeed, a striking performance enlivened throughout by a strong temperament that imparts sparkle even to conventional passages.  But Pareto was right in refusing to sanction a reprint or a second edition.  For, so far as pure theory is concerned, there was nothing specifically Paretian about it.  It was only after 1897 that he rose to heights of his own.  The first major publications that testify to his progress are the Sunto (*Giornale degli Economisti*, 1900) and the *Résumé* of his Paris course.[3]  The *Manuale* or rather, because of the appendix, the *Manuel* (1909), then marks the point of highest elevation that he reached.

The structure of the tower he erected on that spot is far from faultless.  Many things that are essential in a comprehensive treatise received but scant attention.  I do not mean merely that Pareto's work cannot stand comparison with Marshall's in those qualities that are ordinarily looked for in a "manual."  Much more serious is it that important parts of the *theoretical* organon are inadequately thought out.  Pareto's theory of money, for instance, is on the whole inferior to that of Walras.  His theory of capital and interest derives all its merits from that of Walras.  As regards interest he seems to have been content to rely for explanation on the fact that items of physical capital, hence their services, are not free goods.  His theory of monopoly cannot, I believe, be salvaged by even the most generous interpretation.[4]  In spite of all this, the adverse judgment arrived at by some critics is completely wrong.  For it neglects not only many individual strong points but, much more important, the essence of

2.  Walras was perfectly aware of all the short cuts he had had to make in order to finish what he himself realized was a provisory structure.  He never believed that such assumptions as those of constant coefficients of production, of timelessness of production, of absence of overhead costs, of equal size of firms could or should stand forever.  It cannot be averred that, in this respect, Pareto was quite just to him.  It was not only that Walras' was the pioneer performance: Walras had also indicated what was to be done next.

3.  See above, p. 149, note 3.

4.  Some merit there was, however, in his inclusion of the theory of monopoly in the body of general theory.  Also, his theory of international trade must not be reduced, as it mostly is, to a mere criticism of comparative costs.  He sketched, although he did not elaborate, a theory of his own that was the first to apply to international trade the apparatus of general equilibrium.  See v. Haberler, *Theory of International Trade*, 1936, p. 123.

the achievement. The most important of those strong points, the theories of value and of production, will be discussed presently. But first we must try to define that achievement itself of which these two theories were but applications.

The first idea that must occur, from a purely theoretical point of view, to anyone who has mastered Walras' system is to raise it to a still higher level of generality. When we follow Walras and, indeed, all the marginal utility theorists on their progress through the phenomena of exchange, production, and so on, we discover that they are trying to solve problems that in ultimate logic reduce to one only: *all* their problems — not only the problems of production — are problems in the transformation of economic quantitiés and formally alike, the differences consisting merely in the different restrictions to which economic action is subject in different fields. Suppose we decide to do what we do in all the sciences, that is, to separate out the common core of all economic problems and to build a theory of this common core once and for all. The viewpoint of "mental economy" (E. Mach's *Denkökonomie*) will justify this endeavor to utilitarians. A theory of this kind will work with quite general indices, such as "tastes" and "obstacles," and need not stop at the specifically economic meanings that we may assign to these words. We may transcend economics and rise to the conception of a system of undefined "things" that are simply subject to certain restrictions and then try to develop a perfectly general mathematical *logic of systems*. Stretches of this road should be quite familiar to economists who have for generations used primitive devices, such as our venerable friend Crusoe, for the purpose of displaying certain features of economic logic. Pareto simply did the same on a much higher level and on a much broader front. But in these altitudes it is difficult to breathe and still more difficult to gain ground. Critics as competent as the late A. A. Young have been of the opinion that Pareto achieved nothing but "arid generalizations." But only the future can tell whether this is so. Meanwhile we should recognize the greatness of the attempt.

An example will show that such a "rush for generality" may produce not only logical stones but also economic bread, though it suffers from the weakness that it still moves on a relatively low level of generality and, indeed, dates from the *Cours*. As everyone knows, Marx's work is an analysis of the capitalist process, no doubt geared to the purpose of showing that this process will issue in a socialist society but entirely free from any attempt at blocking out the economics of this society. And there are a number of Marxist and

neo-Marxist contributions to the latter problem that can only be described as complete failures. As everyone also knows by now, the service to socialist doctrine that Marxist theorists have been unable to render has been rendered by E. Barone whose famous paper on the subject ("Il Ministro della produzione nello stato colletivista," *Giornale degli Economisti*, 1908) has been surpassed by modern writers only in secondary details. But the essential idea of Barone's argument is clearly indicated in the second volume of Pareto's *Cours* (p. 94) and in his *Manuel* (p. 362), namely, the idea to lift, as it were, the logical core of the economic process above the ground of the institutional garb in which it is given to observation. The reader will observe how easily this idea suggests itself, as a special case, once we place ourselves on the standpoint of Pareto's general theory of tastes and obstacles, although it also suggested itself to Wieser.

In this special case, Pareto has nearly lost his claims to priority — at least among Anglo-American economists — although he had not only posited the problem but also pointed out the way to its solution. In other cases, he lost them completely because he confined himself to mere suggestions. Thus, assisted by hindsight, we may discern in the *Manuel* many pointers toward the later economic dynamics. However, none of them, such as his reference to a form of adaptation similar to a *courbe de poursuite* (dog-and-his-master problem, see, e.g. p. 289) and to the presence of *vibration continuelle* (see, e.g. p. 528), was put to any use other than the negative one of showing that the economic system's tendency toward a unique and stable "solution" (i.e., a unique set of values that will satisfy its conditions) is a much more doubtful matter than the economists of that period, including Walras, supposed.[5] No positive use was made of these suggestions,[6] and no method was indicated for attacking these problems. I therefore think that we should not hesitate to describe Pareto's work as static theory, and that substantial justice is done if we add that he, more than others, was aware of its limitations and of the call of the problems beyond.[7]

5. See, e.g. his discussion of unstable equilibrium in the article in the *Encyclopédie des sciences mathématiques*.

6. The jejune theory of crises (pp. 528–38) certainly does not qualify for being listed as an exception.

7. Pareto himself (p. 147) divided the subject of pure economics into statics; a dynamics that studies successive equilibria and seems to me to denote comparative statics; and another dynamics that studies the *mouvement du phénomène économique* and seems to merge genuine dynamics with the problems of evolution in a manner that would have proved highly inconvenient but for the fact that both remained, with Pareto, quite rudimentary. I know that the situation must look different to a disciple. But although the latter's attitude has its place, it cannot be adopted here.

We now proceed to a brief discussion of Pareto's work in the fields of value and production keeping in mind that, from the standpoint adumbrated above, they really merge into a single theory.

Most modern theorists, although not all, will agree that the historical importance of the utility and marginal utility theory of Jevons, Menger, and Walras rests mainly upon the fact that it served as the ladder by which these economists climbed up to the conception of general economic equilibrium although this conception was much more clearly perceived and much more fully developed by Walras than it was by either the Austrians or Jevons.[8]  In other words, the utility and marginal utility theory was one of several possible avenues to the thing that really mattered and, besides offering an excellent method for demonstrating in an easily understandable way the relations that hold the economic system together and, in fact, make a unified system out of the mass of economic phenomena which departmentalize so easily, had no great importance in itself.  Or, to put it still differently, utility theory was an extremely useful heuristic hypothesis and nothing more.[9]  But neither Walras nor the Austrians were of this opinion.  On the contrary, for them the utility theory was nothing less than ultimate truth, the discovery of the key to all the secrets of pure economics.  In consequence, they placed an emphasis upon it that in turn induced Pareto and the Paretians to

8. As Lord Keynes in his biographical essay on Alfred Marshall has pointed out, Marshall was also in full possession of that conception, and we have Keynes's word, as well as other indications, for believing that he arrived at this conception independently and earlier rather than later than Walras.  This does not alter the fact, however, that he published nothing about it that antedates the relevant notes in the Appendix to the *Principles* (notes XIV and XXI of the fourth edition) which, moreover, cannot be described according to the ordinary rules of assigning priority as more than glimpses. We have, therefore, to conclude that Walras' priority is unchallengeable.  But so is that of the Austrians and especially of Wieser.  It is perfectly clear that it was only lack of mathematical skill and especially the inability to handle systems of simultaneous equations that prevented Menger from producing an exact system that would have been substantially similar to that of Walras.  But I do not think that those historians are right who attribute the concept of general equilibrium already to Cournot. Chapter XI of the *Researches into the Mathematical Principles of the Theory of Wealth* contains nothing but recognition of the general interdependence of economic quantities, and neither there nor anywhere else does Cournot offer guidance toward the great attempt to make this conception explicit and fruitful. All the actual work in the *Researches* is either partial analysis or else, to some extent, aggregative analysis.

9. I wish to make it quite clear, however, first, that I do not think that its heuristic value is as yet exhausted, and, second, that the statement in the text must be read with the proviso, "so far as the purpose of establishing the determinateness and stability of static equilibrium is concerned."  There may be other uses for it and it is impossible to be sure that for such other purposes it might not be revived any moment.

place undue emphasis upon their renunciation of it. Authors in the English-speaking world, particularly Professors Allen and Hicks, followed suit and very generously congratulated Pareto on what to them also seemed a new departure of first-rate importance. In fact, there is a widespread opinion to the effect that this new departure constitutes Pareto's main contribution.

There are indications in the *Cours* that Pareto was not quite satisfied with the Walrasian theory of value from the first. But his amendments, either insignificant or not original, remained within the precincts of the principle itself. Of the insignificant amendments, we merely mention the introduction of the term *ophélimité* in the place of the term utility (*ophélimité élémentaire* for marginal utility or Walras' *rareté*) on the ground that the latter carried too many misleading associations. Of those that are not original with Pareto I mention the conception of utility and marginal utility as functions of all the commodities that the consuming unit possesses or consumes in an appropriately chosen period of time, instead of Walras' conception of total and marginal utility of every commodity as function of the quantity of that commodity alone. This obvious improvement is due to Edgeworth, but I confess to some doubt whether Edgeworth was fully aware of the theoretical difficulties this improvement would cause, for it turns the final degree of utility that was simply an ordinary differential coefficient with Jevons, Walras, and also with Marshall, into a partial differential coefficient and this greatly increases the mathematical difficulties we encounter when trying to prove the determinateness of the economic system even in its most simplified form.[1]

Before long, however, and certainly before 1900, the year in which he delivered his Paris lectures that made his change of standpoint publicly known, Pareto realized that, for his purposes at least, the concept of measurable utility (cardinal utility) might be safely abandoned[2] or that, in any case, it would have to be abandoned for

1. To be more exact: when we are trying to prove that there is one and only one set of values that will satisfy the equations of general equilibrium, so far as I can see everything is plain sailing so long as we adhere to the assumption that the marginal utility of every commodity depends only on the quantity of that commodity and so long as we do not admit any money but only a *numéraire*. The restrictions that are necessary even then in order to produce proof of determinateness seem to me quite bearable economically. It is the intrusion of those partial differential coefficients which creates the real difficulty.

2. Gustav Cassel came to the same conclusion in 1899. He went even further than Pareto and claimed to be able to do without any utility concept at all. It is not possible to explain here why this claim was unjustified, and why his method of starting with market demand curves which he simply postulated is inadmissible. However, in order to appreciate that episode in the history of

reasons that were first stated exactly in the second part of Irving Fisher's *Mathematical Investigations into the Theory of Value and Price* (1892). To save the situation he appealed to the indifference and preference curves that had been first introduced by Edgeworth. But, whereas Edgeworth still started from a measurable total utility from which he deduced the definition of these lines, Pareto inverted the process. He took the indifference lines as the given thing to start from and showed that it was possible to arrive from them at the determination of economic equilibrium in pure competition and also to proceed to certain functions which might be identical with utility if it exists. In any case, it was possible to obtain (ordinal) *indices* of utility or what Pareto called index functions (*Manuale*, p. 540, note 1).

I wish to bring out two points quite clearly. The first is that Pareto, though he may have adapted an invention of Edgeworth's to his own use, imparted to the indifference varieties a meaning that they do not carry in Edgeworth's *Mathematical Psychics*. They are quite divested of any utility connotation, and what the concept of utility had done for the theory of economic equilibrium was now to be done by certain assumptions about the form of these indifference curves. The new idea was to replace utility postulates by postulates about observable behavior and thus to base economic theory on what seemed to Pareto to be more secure foundations. It might be urged of course that in spite of several attempts nobody has as yet succeeded in carrying out such observations and that it is difficult to indulge in the hope that we might construct them from objective data *in their whole extent* so as to derive a complete empirical indifference map. Let us call them, therefore, potentially empirical or, to misuse a Kantian term, "referring to possible experience." In any case, their introduction for a purpose entirely foreign to Edgeworth's might be called a truly original achievement were it not for the fact that this achievement had been, as Pareto recognized, foreshadowed in Fisher's work mentioned above.

The second point is that Pareto's own argument brings out the difficulty he experienced in disentangling himself entirely from the

economic theory it is necessary to remember that at that time not only the utility theory itself but also all the theory of cost and distribution that was built upon it still met with great resistance. This resistance was, especially in Germany and Italy but also elsewhere, sometimes motivated by objections against operating with unmeasurable and unverifiable psychic magnitudes. And so the opposition to marginal utility theory from Pareto and others joined forces with (or at least added new vitality to) a common-run argument that had been repeated again and again by writers with whom Pareto would not have cared to find himself associated.

old utility theory. He always kept an eye on the cases in which it might be possible to speak of utility and even of cardinal utility, the existence of which — hence the question of integrability — continued to interest him very much. And his index functions bear after all a pretty close similarity to the old concept. In fact, as has been pointed out by Allen and Hicks, he never succeeded quite in disentangling himself entirely, and he continued to use concepts such as the Edgeworthian definitions of rivalry and complementarity that do not go well with his fundamental idea. This fundamental idea, let us add, was developed and defended as early as 1902 by P. Boninsegni.[3] By 1908, Enrico Barone, in the paper mentioned already, definitely went beyond Pareto by confining his fundamental assumptions in the matter of value theory to what he called the *fact* that, confronted with given prices of products and productive services, every individual distributes his receipts from the sale of his services between expenditures on consumption goods and saving in a certain unique manner "of which we are not going to investigate the motives." This, so he pointed out, does away with any concept of either utility or indifference functions. The rest of the story is too well known to detain us. I shall merely mention the papers of Johnson and Slutsky that for the time being remained practically unnoticed; the important reformulation of Bowley in his *Groundwork* that was more influential; and the work of Allen and Hicks, Georgescu-Roegen, Samuelson, and H. Wold. If we accept the present situation as "provisionally final," we must indeed salute either Fisher or Pareto as the patron saint of the modern theory of value.

But, still more definitely than patron saint of the modern theory of value is Pareto the patron saint of the "New Welfare Economics." The story of how, once more, he came to render a service to a cause with which he was — or would be — completely out of sympathy, is not without its humor. From the very beginnings of economics, a loosely defined public welfare played a great role in the writings of economists. The familiar slogans of utilitarianism (Beccaria, Bentham) did something toward rationalizing the concept, and the utility theory of value seemed admirably qualified to implement it: in fact it was promptly harnessed to the task, e.g. in the field of taxation. The Fisher-Pareto theory of indifference varieties, destroying as it did the bases of arguments that worked with cardinal utility or even with interpersonal comparison of utility (satisfaction), should, so we might think at first blush, have done away with all this. But instead

3. "I fondamenti dell' economia pura," *Giornale degli Economisti*, February 1902.

of drawing this conclusion — and in spite of his contempt for the political humanitarianism of our age — Pareto immediately went on to attack the problem of maxima of *collective* satisfaction afresh. The definitive formulation was left for Barone,[4] but the main idea is again Pareto's. He observed, first, that all changes imposed upon any given economic pattern may be said to increase welfare or collective satisfaction in a perfectly objective sense if those who gain in terms of *numéraire* could compensate those who lose in terms of *numéraire* and still have some gain left. This criterion will in fact salvage some though not all of the welfare judgments usually passed by economists.[5] Second, Pareto pointed out that welfare judgments that cannot be salvaged in this manner must be explicitly based on extra-economic, e.g. "ethical," considerations. And third, he showed (pp. 363–4) that the criterion may be used in order to establish that *l'état collectiviste* may improve upon the level of welfare that is practically attainable under perfect competition.[6] But, barring developments, these points are pretty much what the New Welfare Economics amounts to.

That part of Pareto's welfare economics which deals with the logic of production provides a convenient transition to his second great contribution to pure theory, his theory of production.[7] Approaching the problem from the side of the theory of choice and applying to the producer's case the general apparatus of indifference curves and derivative concepts (*lignes du plus grand profit, lignes de transformations complètes et incomplètes* etc.), he sketched out a comprehensive structure only parts of which are explicitly present in the literature of his time[8] and which may be said to constitute the

4. See "Il ministro. . ."p. 276, (mentioned above, p. 159).

5. The criterion, in strict logic, is independent of whether that compensation is actually made or not. In the latter case, we simply split the change imposed into two parts: into a change that improves collective satisfaction to which the criterion applies and a transfer from losers to gainers to which it does not apply. Even so, I do not wish to appear in the role of an advocate of the welfare judgments that the criterion protects from being invalidated by objections against the use of cardinal utility or of interpersonal comparison of states of satisfaction. There are other and much more important ones, especially the objection that these "objective" welfare judgments neglect all but the immediate effects.

6. The last sentence on p. 363 of the *Manuel* seems to me to anticipate substantially Professor Hotelling's argument in "General Welfare in Relation to Problems of Taxation and of Railway and Public Utility Rates", *Econometrica*, VI, (1938). The practical application to railroads of the principle that welfare might be maximized by charging prices that will cover marginal cost even in the case of decreasing cost industries and by financing fixed costs (as Pareto said) in some other way, is old. So far as I know, it is due to Launhardt who inferred from this that investment in railroads should "never" be left to private industry (*Mathematische Begründung der Volkswirtschaftslehre*, 1885, p. 294; and earlier writings.)

foundation of the mathematical theory of production of our own age or, at all events, of its statics. In particular, its very generality leaves room for all the special cases that we may wish to treat without placing exclusive emphasis on any one of them: the "obstacles" may be anything at first, and can then assume any of the forms that occur more commonly in practice — the factors that are required in fixed quantities irrespective of output, the factors that are required in technologically determined quantities per unit of output, the "compensatory" factors, and so on, all take their places in a theoretically complete schema of possibilities. In appraising this achievement, we must keep in mind that Pareto was primarily concerned with generalizing and otherwise improving the work of his great predecessor. Again his work may be divided into a first part that culminated in the *Cours* and a second part that culminated in the *Manuel*, though some minor touches were added in the article in the *Encyclopédie des Sciences Mathématiques* (Volume I, 1911).

Originally, Walras had expounded his theory of production on the assumption of fixed coefficients of production — fixed (average) inputs per units of output — not because he believed that this was the only or even a very important case but because he thought himself justified in adopting what he considered to be a simplification.[9] His reply to private criticisms that poured in on him was that "the economists who will come after me are free to insert one by one all the complications they please. They and I, so I think, will then have done everything that it was our duty to do" (*édition définitive*, p. 479). So far as this goes Pareto cannot be said to have done more than take Walras' advice. In addition, when the *Cours* appeared, Walras had already introduced variable coefficients, on a suggestion of Barone's that reached him in 1894,[1] though without altering the argument of

7. See in particular *Manuel*, Chapter III, paragraphs 74–82, 100–105; Chapter V, and paragraphs 77–107 of the Appendix.

8. But if we leave out the word "explicitly," then a much greater part of Pareto's schema must be credited to some of his contemporaries, or even predecessors, and especially to Marshall.

9. It is curious that the greatest of all theorists should have entertained that opinion. For, first, this simplification creates analytic difficulties that may set us wondering whether in the end it is a simplification at all; and, second, it creates a gulf between theory and reality that is great enough to make it doubtful whether results obtained by means of it are of any use.

1. This was done in a *Note* published in 1896 and reprinted in the third edition of the *Eléments*. In the fourth edition (1900) a full-fledged marginal-productivity theory was presented in the 36th *leçon* in a form that was open to criticism on various counts and was later on revised for the final edition that was published posthumously in 1926. See on this and for a useful rendering of Pareto's later theory: H. Schultz, "Marginal Productivity and the General Pricing Process," *Journal of Political Economy*, October 1929.

the fundamental section on production. In the same year (1894) appeared Wicksteed's *Essay on the Coordination of the Laws of Distribution*. Finally, variable coefficients of production were no novelty in any case after all that Jevons, Menger, and Marshall had said on the subject. Pareto's *Cours* added only an elegant formulation and a number of reasons — not all of them convincing — why the case of compensatory coefficients should not be treated as the only or the fundamental one either.

It is of course a mere matter of terminological taste whether or not we are to confine the phrase "marginal productivity theory" to this case.[2] Pareto did so confine it and, in the years following upon the publication of the *Cours*, grew increasingly hostile to it, declaring it definitely "erroneous." He was evidently under the impression that he had refuted or, at any rate, outgrown it in the same sense in which he felt that he had refuted or outgrown the marginal utility theory. His brilliant theory of cost — which, among other things, withdraws from their dangerously exposed positions the textbook theorems that, in perfect equilibrium of pure competition, price should equal marginal cost and total receipts should at the same time equal total cost — permit us to test this claim.[3] So far as productive combinations depend on economic considerations — and, after all, it is the economic considerations which it is the economists' task to clear up — the difference, as compared with straight marginal productivity theory, is not great. But Pareto does teach us how to handle the deviations from it that technological and social restrictions impose. And, here as elsewhere, he does something else: he always points beyond himself.

2. The main reason for doing so is a textbook tradition which takes into account only production functions that represent quantities of product as dependent on "substitutional factors" alone and arrive at the theorem that, in perfect equilibrium of pure competition, the unit of each of the innumerable requisites of production earns a compensation that equals physical marginal productivity times the price of the product. But we do not leave the precincts of the marginal productivity argument if we admit "limitational factors" or, more generally, restrictions upon the production function that will produce results at variance with that theorem. See, e.g. A. Smithies, "The Boundaries of the Production Function and the Utility Function," *Explorations in Economics, Notes and Essays contributed in Honor of F. W. Taussig*, 1936.

3. We use this opportunity in order to advert to Pareto's conception of rent which arises in the cases where those two conditions (total cost = total receipts; price = marginal cost) are incompatible, and especially in the cases where transformation of savings into certain kinds of capital goods meets with difficulties. This theory of rent has experienced a renaissance in our days. It may help us toward an improved theory of friction. But it can hardly do more.

## III. THE SOCIOLOGIST

There is nothing surprising in the habit of economists to invade the sociological field. A large part of their work — practically the whole of what they have to say on institutions and on the forces that shape economic behavior — inevitably overlaps the sociologist's preserves. In consequence, a no-man's land or everyman's land has developed that might conveniently be called economic sociology. More or less important elements that hail from that land are to be found in practically every economic treatise or textbook. But beyond this many economists, and especially those who define economics proper rather strictly, have done sociological work. A. Smith's *Moral Sentiments* and Wieser's *Gesetz der Macht*, are both outstanding instances of a large genus. But few if any men in the list of great economists have devoted so large a part of their energy as has Pareto to what at first sight seems to be an extra-curricular activity, and few if any owe so much of their international reputation to what they have done in that field. But his achievement is not easy to characterize and to appraise. The enthusiastic applause of some and the hostility of others are both understandable but neither can be taken quite seriously because the non-scientific sources of both are painfully obvious in most cases. Although several minor works and a large number of newspaper articles would have to be considered in order to give a satisfactory picture, we need not go beyond the *Systèmes Socialistes*, the *Manuel* (especially Chapters 2 and 7) and the *Trattato di Sociologia Generale*.

Let us begin with two aspects of Pareto's sociology that are perfectly obvious and the reverse of difficult to characterize. First, although Pareto the economist touched upon a large number of extremely concrete and practical problems throughout his long life, his purely scientific contribution is in the realm of the most abstract economic logic. It is, therefore, quite understandable that he should have experienced a wish and, in fact, a need to erect alongside his pure theory another building that would shelter facts and reasonings of a different kind, facts and reasonings that would do something toward answering the question how the elements taken care of by his economic theory might be expected to work out in practical life. Second, we have seen that in his earlier days, at least as long as he lived in Italy, he had taken a passionate interest in the debates on questions of economic and general policy. The born thinker that he was must have been struck by the impotence of the rational argument, and the question must have intruded upon him of what it really is

that determines political action and the fate of states and civiliza-
tions.   Again, it is quite understandable that, so soon as he had
settled down to a life of thought, this question should have emerged
from the sphere of easy and superficial answers that all of us are prone
to give when immersed in our daily work, and that he should have
attempted to raise it to the plane of scientific analysis.  This amounts
to saying that primarily and fundamentally his sociology was a
sociology of the political process.  Of course, everything that man
does or thinks or feels and all his cultural creations and his attitudes
toward cultural creations are bound to come in somehow or other
when we think about the political process which then becomes but a
special case.  But it was this special case which fascinated Pareto and
for the sake of which he erected and adorned a much larger structure.

Next still moving on ground that is relatively easy to survey, we
shall consider his method.   Pareto himself emphasized again and
again that he simply applied the same "logico-experimental" methods
that had served him for the purposes of economic theory to the task
of analyzing the "experimentally" verifiable reality of other aspects
of social life, allowing himself to be guided here as elsewhere by the
example of the physical sciences.   This was, of course, a complete
delusion. It is easy to observe, for instance, that he made large and
in part illegitimate use of psychological interpretations for which
there is no analogy in the physical sciences and that his material,
such as it was, was the product of observation and not of experi-
ment — a difference which is fundamental from the standpoint of
method.  I am afraid that what he really meant to emphasize when
trying to formulate his rules of procedure was simply the detachment
of the philosopher who does not identify himself with any party,
interest, or creed.   The possibility of such detachment raises, of
course, a very well-known fundamental difficulty and one that Pareto
was the less qualified to overcome because he failed to see it. Actually
he used two different analytic schemata: one that may be called a
morphology of society and does invite the use of facts that are, poten-
tially at least, amenable to observation in a similar sense as are the
facts of anatomy or biology; and another that pertains to social
psychology.   Both schemata are indeed *illustrated* or even, to some
extent, *verified* by historical and contemporaneous instances, but
neither is *derived* from them by anything like a "logico-experimental"
method: both are reflexes of a highly personal vision of the social
process that owes much to Pareto's background, practical experi-
ence — and resentments.   The affinity of the morphological schema
with Darwinian selection and of the socio-psychological schema with

parts of the teaching of Tarde, Dürkheim, Lévy-Bruhl, and Th. Ribto is obvious. Still more so is the relation of both with the current of thought glanced at in the first section of this essay that issued in derogatory criticism of the doings of parliamentary democracy — the current that was anti-intellectualist, anti-utilitarian, anti-equalitarian and, *in the special sense defined by these terms*,[4] anti-liberal. But the force of the man created from these materials something that was nevertheless specifically his own.[5]

The morphological schema centers in the proposition that all societies consist of heterogeneous masses of members — individuals or families — and are structured according to the aptitudes of these members for the relevant social function: in a society of thieves, the *ex hypothesi* widely varying ability to steal would determine social rank, and hence influence upon the government of the society. Pareto seems to assume that these abilities, while capable of improvement and of decay, are substantially innate though he makes little effort to establish this. Moreover, though distributed continuously in the population, they lead to the formation of classes, the "higher" ones of which have and use the means of buttressing their position and of separating themselves from the lower strata. In consequence, there is in the lowest strata a tendency to accumulate superior ability that is prevented from rising, and in the topmost stratum, in the aristocracy or *élite*, a tendency to decumulate energy through disuse — with resulting tension and ultimate replacement of the ruling minority by another ruling minority that is drawn from the superior elements in the *couches inférieures*. This *circulation des élites* does not, however, affect the principle that it is always *some* minority which rules, and does not do anything to bring any given society nearer to the ideal of equality, though it does produce equalitarian philosophies or slogans in the course of the struggles that ensue. With a turn of phrase that recalls the first sentence of the *Communist Manifesto*, Pareto proclaimed that history is essentially a history *de la succession des aristocracies (Manuel*, p. 425). But his presentation of this part of his argument is so very sketchy and he leaves his readers with so much to

4. This proviso is very necessary. There are other meanings of the word "liberal," one of which would describe Pareto's position much better than could any other term. Similarly, there is a sense in which he might be justly called a great humanitarian. But it is not the one which he applied to *individus dégénérés, d'intelligence et de volonté faibles (Manuel*, p. 130.)

5. It is highly instructive to observe how different the results are that different men arrive at not only from the same facts but also from the same intuitions. Graham Wallas was an orthodox English radical and a Fabian. But in *Human Nature in Politics* he drew a picture that was not in the least more flattering to the slogans of political democracy than was Pareto's.

interpolate that I am not at all sure that I have rendered justice to his thought. Nevertheless, I had to make the attempt. For some such argument is necessary in order to put his social psychology into its proper light.

The socio-psychological schema centers in the concept of the non-logical (not necessarily illogical) action. This concept recognizes the well-known fact — well-known, in particular, to economists — that the great mass of our everyday actions is not the result of rational reasoning on rationally performed observations, but simply of habit, impulse, sense of duty, imitation and so on, although many of them admit of satisfactory rationalization *ex post* either by the observer or the actor. So far there is nothing in Pareto's psycho-sociology that could be unfamiliar to anyone. What is unfamiliar however is his tremendous emphasis upon the additional facts that a great number of actions — and let us add at once, beliefs — are being rationalized, both by actors and by observers, in ways that will not stand up under scientific analysis and, more important, that some actions and beliefs are altogether incapable of being rationalized in any way that will. The importance of this second step for a sociology of the political process becomes obvious if we take a third one: Pareto maintained that the large majority of all the actions and beliefs that make up that process are of the type mentioned last. Take, as an instance on which we all agree, the idea of the Social Compact or, as an instance on which most of us agree, Rousseau's theory of the *volonté générale*. Only, according to Pareto, practically all the actions, principles, beliefs and so on prevailing in the collective mind of electorates belong in the same category. And a large part of the *Trattato* consists in illustrating this, often amusingly, sometimes instructively.

It will serve our purpose to put this point strongly, more strongly than Pareto himself ever put it. The masses of thought and the conceptual structures that form the conscious surface of the social and in particular of the political process have no empirical validity whatever. They work with entities such as liberty, democracy, equality, that are as imaginary as were the gods and goddesses who fought for and against Greeks and Trojans in the *Iliad*, and are connected by reasonings that habitually violate the rules of logic. In other words, from a logical standpoint, they are nonsense unalloyed. This makes a political philosophy that is best described by its diametrical opposition to that of Jeremy Bentham. It should be observed, however, that this diagnosis of the political myths (Sorel) did not induce Pareto to overlook the function that this logical nonsense may fill in national

life. After having gone through with an analysis that is severely positivist in nature, he refused to draw the conclusion that would seem the obvious one to the positivist. While political creeds and social religions — with Pareto there is very little difference between these two — contribute to dissolution in dissolving civilizations, they also contribute to effective organization and action in vital civilizations. This is a very curious attitude for a thorough-going positivist to take and will perhaps be cited at some future time as an outstanding example of the mentality of an epoch that destroyed one type of metaphysical beliefs while ushering in another. It reminds me of the advice which I have heard some psychoanalysts give to some of their patients, namely the advice to cultivate with a view to possible remedial effects a sort of synthetic belief in God. There is of course no contradiction between maintaining that social and political creeds have no empirical significance and admitting that some of them may make for social cohesion and efficiency. But the social philosopher who should thereupon undertake to advise the adoption of the latter would run into the same difficulty as our psychoanalyst: so long as his *analysis* is being accepted his advice must be ineffective, for no synthetic God can be trusted to help; so soon as his *advice* is accepted his analysis will have to be rejected.

That tissue of creations of our imagination Pareto called *dérivations*. The argument adumbrated in the preceding paragraph abundantly shows that they are not without importance as factors that help to shape the historical process. It was Pareto's opinion, however, that this importance is relatively small and that substantially these *dérivations* do no more than verbalize something more fundamental that comes much nearer to determining actual political behavior and the sum total of non-logical actions. Now if we defined this more fundamental something in terms of group interests, and if we then went on to define these group interests in terms of the social location of groups within a society's productive organization, we should be, to say the least, very near Karl Marx's view of the matter, and there is in this point actually a strong affinity which I think it important to emphasize. In fact, if we adopted this line of reasoning, there would be only two major points of difference left between Marxian and Paretian political sociology. On the one hand, Pareto introduced explicitly an element that is only implicitly present in the Marxist analysis: the importance for the explanation of an actual stretch of history, of the greater or smaller degree of social flexibility that a given society displays, or, in other words, the importance of the fact that there exists an optimum of vertical mobility and of resistance

to it that will better than others guarantee what might be termed stability of political change. On the other hand, we need only recall our sketch of Pareto's social morphology in order to realize that with Pareto the historical process is not so much the result of the conflict of comprehensive social classes as it is the result of the conflict of their ruling minorities. It is submitted that, while both differences are to the credit of Paretian sociology, they do not amount to more than corrective improvements upon the Marxist schema. I might add the fact that property relations *per se* are much less in evidence with Pareto than they are with Marx, and that this also constitutes a claim to superiority of the Paretian analysis. But it will be readily seen, that this point is really implied in the other two.

Actually, however, Pareto did not follow up this line of analysis. With him the link between the tissue of delusions which he called *dérivations* and the objective determinants of actual behavior was supplied by what he called the *résidus*. I am conscious of the danger of being unfair if, for the sake of brevity, I define these *résidus* as impulses generally found to be present with human beings that revive, and not in a very inviting manner, the old psychology of "instincts." We need not discuss the list that Pareto drew up — and which contains such items as an instinct of combinations, the sexual impulse, and so on — especially as Pareto himself does not seem to have been very satisfied with it. It is sufficient to point out the obvious methodological objection to any such procedure; even if Pareto's *résidus* and the "laws" of their association and persistence were much more satisfactorily analyzed than they are, they would still be labels rather than solutions of problems, and call for professional investigation of a kind for which Pareto lacked the equipment. It is therefore quite understandable that Pareto's work has exerted so little influence upon professional sociology and social psychology, and that professional sociologists and social-psychologists have but rarely displayed a sense of the greatness of the structure as a whole.[6]

But those and other shortcomings are not decisive. Pareto's work is more than a research program. Also, it is more than mere analysis. The fundamental principle that what individuals, groups, and nations actually do must find its explanation in something much deeper than the creeds and slogans that are used in order to verbalize action, conveys a lesson of which modern men — and none more than we economists — stand much in need. We are in the habit when discussing questions of policy of accepting at face value the

6. Professor Talcott Parsons' analysis of Paretian sociology stands almost alone in the Anglo-American sociological literature.

slogans of our own and, indeed, of a by-gone time. We reason exactly as if the Benthamite creed of the eighteenth century had ever been valid. We refuse to realize that policies are politics and to admit to ourselves what politics are. We cultivate the subnormal and do our best to suppress whatever there is of strength and sparkle. In conditions such as these, Pareto's message, however one-sided, is a healthy antidote. It is not, like his economics, a technical achievement of the first order. It is something quite different. It is an attempt to preach a sermon.

<div align="right">JOSEPH A. SCHUMPETER.</div>

HARVARD UNIVERSITY

# THE ACCELERATOR
# AS A GENERATOR OF STEADY GROWTH

## SUMMARY

Introduction: The theory of the accelerator to date, 174. — How steady growth is generated, 176. — Influence of the investment period, 180. — Influence of the income period, 181. — A more general case, 182. — Qualifications, 186. — Price as a stabilizer, 188. — Appendix, 193.

The early development of the theory of the accelerator by such economists as Aftalion[1] and Clark[2] stressed its role as a generator of cycles. Samuelson, however, pointed out that the interaction of the accelerator and the propensity to consume can lead to a variety of time patterns of the national income depending on the numerical values of these factors.[3] Among the possible patterns was one of steady growth for large values of accelerator and propensity.

Meanwhile Domar[4] and Harrod[5] investigated the rate of growth of income required to induce (or justify) enough investment expenditure to offset the savings that would be associated with the growing income, so that the growth of income could be maintained. Schelling has concluded that their quest was hopeless — "that growth via induced invest· ·n , is unreliable, that a stable dynamic equilibrium cannot be go· _ :·a acceleration."[6]

Schelling drew this conclusion from a very restricted set of examples. It is the purpose of this paper to examine the problem in greater generality. The conclusion here reached is that plausible values of accelerator and propensity to consume can indeed permit steady growth, but with constant prices they will lead to so rapid a rate of growth as to suggest that the assumed relationships cannot long persist. Although there are some numerical values of accelerator and propensity that will yield moderate steady growth, those values

1. Albert Aftalion, *Les crises périodiques de surproduction*, Paris, 1913.
2. J. M. Clark, "Business Acceleration and the Law of Demand," *Journal of Political Economy*, March 1917.
8. Paul Samuelson, "Interactions between the Multiplier Analysis and the Principle of Acceleration," *Review of Economic Statistics*, XXI, May 1939.
4. Evsey D. Domar, "Expansion and Employment," *American Economic Review*, XXXVII, March 1947.
5. Roy F. Harrod, "An Essay in Dynamic Theory," *Economic Journal*, XLIX, March 1939.
6. T. C. Schelling, "Capital Growth and Equilibrium," *American Economic Review*, December 1947, p. 876.

do not seem consistent with what we know about the facts of our economic system.[7]

Furthermore, a given set of values of propensity and accelerator can lead to radically different time patterns of the national income depending upon the periods over which accelerator and propensity operate. Two periods are of fundamental importance in this connection: the time it takes that part of a given income which is spent on consumption to go through the economy and once again become income (the income period); and the time it takes for the investment expenditures induced by a given increase in national income to be made (the investment period).[8] If these periods are short enough there will be rapid steady growth irrespective of the given values of accelerator and propensity to consume so long as the accelerator is not zero.[9] If the income period is short, and the investment period very long, a steady approach to equilibrium is produced. For intermediate values of the time periods cyclical movements predominate.

Gently rising prices may act to stabilize a moderate rate of growth that would otherwise develop into an excessive one. Values of accelerator and propensity and their periods which would lead to tremendous rates of growth of income at constant prices if production could possibly grow that fast can lead to moderate rates of growth of production and of prices if there is a tendency for increased expenditure to raise prices as well as to call forth increased production. Moderate steady growth of the national income  ̄ ⁿrdingly may be generated by plausible values of accelerator and : ⱼₜ ₁sity provided price rises act as stabilizers. There is a very narrow range of price rises which can moderate a rapid rate of growth to a gentle rate of growth. Price rises exceeding this range will lead to cyclical fluctuation even when steady growth at a very rapid rate would take place if prices did not change.

The stabilizing effect of a price rise results from the fact that rising prices reduce the increase of real income associated with a given increase of expenditure, and so reduce the amount of investment

7. In the following discussion the terms "moderate" and "large rate of growth," and "propensity to consume close to one" will frequently be used. They will usually be used in a relative sense, but in order to make the ideas concrete we may adopt the convention that a rate of steady growth of 10% a year or more is "large" rather than "moderate"; a propensity to consume of .95 or greater is "close to unity," and one greater than .98 is "very close to unity."

8. In this article it is assumed that the time it takes expenditure on investment to become income is equal to the income-period. This assumption is dropped in the general case considered in the appendix.

9. As a limiting case if investment depends on the (instantaneous) rate of growth of income, steady growth will be possible irrespective of the value of the accelerator, so long as it is positive.

that will be induced through the accelerator. If the rising prices induce speculative investment, however, the stabilizing effect will be impaired.

## How Steady Growth is Generated

Of course, steady growth will always be possible if the marginal propensity to consume is greater than unity, irrespective of the (positive) value of the accelerator. But such a condition is implausible, and we shall henceforth assume that the marginal propensity to consume is less than unity.

A simple example may demonstrate how an accelerator can introduce steady growth even if the marginal propensity to consume is less than unity. Suppose the marginal propensity to consume is .95, and that it takes a year for a given dollar of income to go through consumption and once again become ninety-five cents of income. That is, income of period $t$ from consumption expenditure $= .95 \, Y_{t-1}$, where $Y_{t-1}$ is total income of period $t$-1. If income is to grow at a steady rate, let us say at 5% a year, each year's income from investment expenditure must equal 10% of the previous year's total income in order that the 5% savings may be offset and 5% increase in expenditure may be supplied. Then the ratio of income from investment expenditure to increase of annual income must be 10% to 5% or 2.0 (the accelerator).

Income of period $t$ equals the sum of consumption and investment expenditures which become income in that period, so that

$$Y_t = .95 \, Y_{t-1} + 2.0 \, (Y_t - Y_{t-1}). \qquad (1)$$

Steady growth will be possible at a rate of 5% per year as may easily be verified by trying $Y_{t-1} = 100$, $Y_t = 105$ in equation (1).

As Schelling has observed, this situation is unstable. Stable equilibrium within any period requires that at higher income expenditures will be less than income, and at lower incomes expenditures will exceed income. This is exactly analogous to the familiar stability condition in supply and demand analysis that the supply curve cut the demand curve from below. Suppose we try $Y_t = 110$, $Y_{t-1} = 100$ in the right-hand side of equation (1). Then the indicated expenditure would be 115, or greater than the assumed $Y_t$. Equation (1) accordingly represents an unstable situation. If equation (1) held in fact, there is no reason to believe that income would grow at 5% a year even though $Y_t = 1.05 \, Y_{t-1}$ is a solution of equation (1). The conditions which underlie equation (1) would lead to an indefinitely large or indefinitely small income. This should be intuitively obvious

from the fact that whenever income is increased by $1, investment will be increased by $2. Since investment is itself part of income, the latter will increase by $2 more than the original $1. This in turn will induce an increase of investment of $4, which, upon immediately becoming income, would in turn lead to an investment increase of $8, and so on *ad infinitum*, all within one time period.

It seems more realistic to assume, then, that income from investment in the current period depends on the change in income of the previous period so that income of period $t$ from investment expenditure $= B \ (Y_{t-1} - Y_{t-2})$
where $B$ is the accelerator. Then, still assuming a propensity to consume of .95:

$$Y_t = .95 \ Y_{t-1} + B \ (Y_{t-1} - Y_{t-2}) \tag{2}$$

If this equation is to yield steady growth at 5% a year, it must be satisfied by the sequence 100, 105, 110.25, for $Y_{t-2}, Y_{t-1}, Y_t,$ respectively. Substituting these values in (2) we find that $B$ must equal 2.1 in order to yield a 5% rate of growth. That is:

$$Y_t = .95 \ Y_{t-1} + 2.1 \ (Y_{t-1} - Y_{t-2}) \tag{2a}$$

yields a 5% rate of growth. Equations (2) and (2a) are stable in the intra-period sense that an increase in the income of period $t$ will not increase expenditure which becomes income in period $t$ at all, since expenditure which becomes income in period $t$ now depends entirely on the incomes of previous periods.

It now seems as though we have constructed an example which yields a moderate rate of steady growth without the intra-period instability characteristic of the first example. If the rate of 5% per year is deemed excessive, we can by choosing a larger value of $B$ construct an example with as small a rate of growth as we like. If, for example, we set $B = 3.57$ in (2) the resulting equation will fit a 2% rate of annual growth. The lower the rate of steady growth, the greater must be the accelerator necessary to induce investment expenditure to more than offset the assumed marginal rate of saving (5%), and thus support the steady growth. Thus with steady growth of 1% a year, and the propensity to consume, $a = .95$, the accelerator would have to be approximately 6, since it must provide expenditure to make up for 5% saving plus 1% growth in income. Similarly, if the growth of income were to be ½% a year (with propensity still .95) the accelerator would have to be about 11.

Yet when the accelerator is large, a small increase in income should induce a large investment expenditure, which, upon becoming

income should induce further large investment expenditure, and so on. Paradoxically then, we should expect a large accelerator to lead to a rapid rate of growth of income, while in the preceding paragraph we saw that a larger accelerator is associated with a smaller rate of growth. Both sides of this paradox are true however, because with each accelerator there will be associated not one, but two, rates of growth.

Equation (2a) is satisfied not only by a 5% increase per period, but also by a 100% increase per period, as may be verified by trying out the sequence 100, 200, 400 in equation (2a). More accurately, equation (2a) will be satisfied by a sequence of incomes made up of two components.[1] One of the two components of the solution of equation (2a) will grow (or decline) at 5% a year and the other at 100% a year. That is, equation (2a) will be satisfied by

$$Y_t = K_1 \,(1.05)^t + K_2 \,(2.00)^t \qquad\qquad (2b)$$

where $K_1$ and $K_2$ are constants which depend only on the incomes which start off the series.[2]

If the initiating series is growing at a rate only slightly over 5%, income will at first grow at a rate close to 5% a year, but will fairly rapidly approach a rate of growth close to 100% a year. These rates of growth need not apply to income measured from zero, but from that level of income at which savings are zero. Suppose savings are zero in the United States at a national income of $50 billion per year. Then $Y_t$ may represent national income over 50 billion dollars. Then $Y_0 = 1$, $Y_1 = 2$ will, in equation (2a), lead to a 100% growth of $Y_t$ per year. This corresponds to an income sequence of 51, 52, 54, 58, 66, 82, 114. Thus a rate of growth of $Y_t$ by 100% a year need not be implausible for a limited number of years.

We may call the two rates of growth the dominant and the minor rate respectively. Let us see whether plausible values of the accelerator cannot yield more moderate dominant rates than in the example given: we shall still assume, however, that $a = .95$.

As $B$ decreases, the minor rate increases and the dominant rate declines, until we consider $B$ smaller than 1.497 at which point all possibility of steady growth disappears and the dominant component of $Y_t$ becomes cyclical. The corresponding minimum dominant rate of growth is over 22% a year.

1. There are two lags, and the general solution of such a difference equation has as many components as the equation has lags.
2. The terms 1.05 and 2.00 may be called growth factors. The relation between a growth factor $(r)$ and a percentage rate of growth per period $(p)$, is then $p = 100(r - 1)\%$.

If the propensity to consume, $a$, in equation (2) were closer to unity than is .95, a smaller dominant rate of steady growth would be possible. For any given $a$ the smallest possible dominant steady growth factor will be $1 + \sqrt{1 - a}$, so that a dominant steady growth factor of 1.05 or less per year would be possible only with a marginal propensity to consume of .9975 or greater under the conditions assumed.

Consequently, under the assumed conditions, values of $a$ which are at all different from unity permit only extremely rapid steady dominant rates. Values of the accelerator that do not generate such high dominant rates will lead to cycles rather than to steady growth. For very small values of the accelerator $B$, the cycles disappear, and only a steady approach toward zero is possible for each component of $Y_t$.

Harrod[3] pointed out the possibility of the existence of a "warranted" rate of growth such that if precisely maintained initially it would induce a continuation of the same rate. But if there is a small displacement from this warranted rate of growth, increasingly large displacements will be generated. This is characteristic of a minor rate of steady growth such as the 5% rate which satisfies equation (2a). If the initial $Y$'s that start off the series are growing at just barely over 5%, then $K_2$ in (2b) will be very small but positive, and the rate of growth of income will at first be close to 5%; but eventually the growth factor of 2.0 will come to dominate that of 1.05 and growth will approximate 100% per year. Thus if we start off equation (2a) with $Y_0 = 100$, $Y_1 = 106$, then $Y_2 = 113.3$, $Y_3 = 123$; but $Y_{10} = 1239$, and $Y_{11} = 2329$. If, however, the initial series is growing at a rate just below 5% a year, $K_2$ will be small but negative, and the rate of growth will steadily diminish and will eventually reach a very rapid rate of decline.

Thus the minor rate of growth is a critical dividing line between initial rates that will lead to an ever-increasing rate of growth approaching the dominant growth factor, and initial rates that will lead to declining rates of growth, and eventually to rapid decline. The frequently observed tendency for growth to be cumulative holds only for rates between minor and dominant rates. Within this range, the more investment there is the faster income will grow and the more investment will be required. Outside the range between

3. Harrod, *op. cit.* Harrod's model differs in essential respects from the model here used. The difference will be discussed in a forthcoming article by the author; it suffices to state here that in Mr. Harrod's model steady growth is always possible and it is always unstable.

minor and dominant rates, a given rate of growth or decline cannot sustain itself.

Earlier investigators of these relationships have been concerned principally with the minor rate of growth, the dominant rate having escaped their attention just as it escaped ours in equation (1). This oversight is the consequence of an assumption of instantaneity of expenditure's becoming income (a zero income period). It will be shown below that for conditions that can yield steady growth, as the income or investment period is shortened, the dominant rate of growth increases without limit. So, as the income or investment period approaches zero the dominant rate becomes infinitely great. But if either period is assumed to be zero and completely ignored, the infinitely large dominant rate of growth is lost from sight. Only the minor rate of growth will then be noticed. The recognition of the instability of the rate of growth thus discovered is an implicit reflection of the operation of the (hidden) infinitely large rate of growth. We may conclude that it is highly misleading to assume that expenditure instantaneously becomes income. The income period is an important characteristic of our economic system which must not be assumed away.

### INFLUENCE OF THE INVESTMENT PERIOD

Values of the ultimate annual accelerator,[4] $B$, that will lead to steady growth are by no means unrealistic, even though the rates of growth they imply could presumably not long be maintained. Certainly the long-run increase of capital may be at least 1.5 to 2.1 times the long-run increase of annual national income. What is unrealistic is that all the investment induced by an increase in the national income should be achieved in a single year following the increase in national income. A slower process is likely, so that a given year's increase in income may initiate a series of investments which will be spread over several years.

We may assume, for the purpose of illustration, that the investment induced by a $1 increase in the annual national income is $2.1 spread evenly over five years, so that a dollar increase of annual income leads to investment of $.42 per year for five years. Then the income in a year $t$ from investment expenditure induced by 5 previous years' increases of income will be:

$$.42(Y_{t-1} - Y_{t-2}) + .42(Y_{t-2} - Y_{t-3}) + \ldots + .42(Y_{t-5} - Y_{t-6})$$

4. The *ultimate annual accelerator* (henceforth referred to as the ultimate accelerator, $B$) is defined as the amount of investment that will ultimately be induced by a permanent increase of $1 in the *annual* national income.

But each intermediate $Y$ occurs once positively and once negatively so that the intermediate $Y$'s cancel out and the income in year $t$ from investment expenditure $= .42(Y_{t-1} - Y_{t-6})$

Let us call the factor .42 the current accelerator $c$, as distinguished from $B$, the ultimate annual accelerator, which is here assumed to be 2.1. If as before $a = .95$ and there is a one-year income period, then:

$$Y_t = .95 \, Y_{t-1} + .42(Y_{t-1} - Y_{t-6}) \tag{3}$$

In this example the dominant rate of growth will be about 20% per year in contrast to the dominant rate of 100% per year generated by the same propensity to consume and ultimate accelerator, when the latter operated entirely within one year instead of over five years. When the accelerator operates gradually over a long period, rather than crowding a great mass of investment into a single year, the dominant rate of growth is reduced. If the ultimate accelerator of 2.1 operated over 6 years (with $a = .95$) the dominant rate of growth would be about 14%; if it operated over 7 years, however, no steady growth at all would be generated, but there would be a cycle of growing amplitude.

So far, our attempts to construct an example of the generation of moderate steady growth by the accelerator have been fruitless. We constructed one that would yield a moderate rate of 5% a year, and found that the 5% rate was dominated by a component that increased at 100% a year. Reducing the value of the accelerator decreased the dominant rate of steady growth but only down to a lower limit set by the value of the marginal propensity to consume. Lengthening of the investment period also reduced the dominant rate of steady growth, but this too switched over to the generation of cycles before a moderate dominant steady growth was achieved, so long as $a$ was not very close to unity.

<center>INFLUENCE OF THE INCOME PERIOD</center>

An income period shorter than one year must also be considered. Thus suppose that the income period is 4 months,[5] but as before $a = .95$, and $B = 2.1$ spread over 5 years. If now we change our unit of time to 4 months, this will change the appropriate value of

5. Cf. Richard M. Goodwin, "The Multiplier," in Seymour Harris, ed., *The New Economics* (New York, 1947), p. 488 for an argument that the income period is probably not longer than ⅓ year. Cf. also Lloyd A. Metzler, "Three Lags in the Circular Flow of Income" in *Income, Employment and Public Policy, Essays in Honor of Alvin H. Hansen,* (New York, 1948). The income period in this article corresponds to the sum of the first two lags considered by Metzler.

the accelerator. An increase of income by $1.00 per third of a year is equal to an increase of $3.00 per year. Therefore, if $B = 2.1$ then $B_{4\,\text{mos.}} = 6.3$. However, the current accelerator $c$, which is defined as the ultimate induced investment divided by the investment period, is invariant with respect to the time unit. An investment period of 5 years will be 15 units of 4 months each, and $c = \dfrac{2.1}{5} = \dfrac{6.3}{15} = .42$.

Therefore the above assumptions are expressed in the equation:

$$Y_t = .95\,Y_{t-1} + .42(Y_{t-1} - Y_{t-16}) \tag{4}$$

The dominant rate of growth indicated by (4) is greater than 36.5% per income period or over 150% per year! If in (4) we try smaller values of $a$, we find that the minimum dominant rate possible so long as the other conditions remain unchanged is 41% per year with $a$ about .78. Smaller values of $a$ will then lead to a cyclical dominant component.

So long as $(a + c) > 1$ the shorter the income period the greater will be the dominant rate of steady growth. This will be shown in the more general case which follows.

A More General Case

We may now investigate the general case of which the preceding examples were special illustrations. We continue to assume that the timing of the propensity to consume can be expressed in the income period; and that the rate of investment expenditure is uniform over the investment period; and that it takes just one income period for any given investment expenditure to become income. Four constants then specify the time shape of the movement of income:

$a$ = the marginal propensity to consume;

$c$ = the current accelerator defined as the ultimate annual accelerator $B$, divided by the investment period in years;[a]

$n$ = the ratio of the investment period to the income period, or the number of income turnovers per investment period;

$m$ = the number of income periods per calendar year.

If we choose the income period as our time unit $\left(1 \text{ unit} = \dfrac{1}{m}\text{th year}\right)$ we need only three of these constants in the equation embodying our assumptions:

$$Y_t = aY_{t-1} + c\,(Y_{t-1} - Y_{t-n-1}) = (a + c)\,Y_{t-1} - cY_{t-n-1} \tag{5}$$

6. Note that $B$ is defined as the increase of capital ultimately induced by a $1 increase in *annual* income.

This equation indicates that each period's income will be $(a + c)$ times the preceding period's income minus $c$ times the income of a much earlier period. This may be conceived as a growth factor $(a + c)$ compounded each income period, and reduced by a factor which is independent of the length of the income period. As the income period shortens, the number of income turnovers per investment period increases, giving $(a + c)$ greater opportunity to compound over the investment period.

If we represent a given investment period as $n$ income periods, $n$ gets larger and larger as the income period gets shorter and shorter, and the growth factor over the investment period approaches $(a + c)^n$ which itself approaches either zero or infinity as $n$ increases, depending, respectively, on whether $(a + c)$ is less than or greater than unity. So a shorter income period leads to a greater rate of growth so long as $a + c > 1$; and it leads to a more rapid approach to an equilibrium value if $a + c < 1$.

Equation (5) will be satisfied by a $Y_t$ which is the sum of $n + 1$ components, of which at most two can grow steadily; the remaining $n - 1$ components will oscillate. The conditions that there be at least one steadily growing component are:[7]

$$\frac{n}{n+1}\,(a + c) > 1 \quad \text{(I)}, \quad \text{and} \quad nc \leqslant \left[\frac{n}{n+1}\,(a + c)\right]^{n+1} \quad \text{(II)}$$

These conditions (provided $a < 1$) are also sufficient to insure that there will be *two* steadily growing components and that each of them will dominate the oscillating components. If $r$ is the growth factor of the dominant steady growth component, then in the absence of outside disturbance, as time goes on, $Y_t/Y_{t-1}$ will in general approach $r$.[8] If conditions (I) and (II) hold, it will also be true that

$$\frac{n}{n+1}\,(a + c) \leqslant r < a + c \quad \text{(III)}, \quad \text{and} \quad nc \leqslant r^{n+1} \quad \text{(IV)}$$

From these four conditions we may draw the following conclusions:

1. *If steady growth is possible at all, the dominant growth factor over the income-investment period*[9] *will equal or exceed the ultimate*

7. See Appendix for proof of these conditions and conclusions.
8. If there are several different dominant growth or cyclical factors with equal absolute values, $Y_t/Y_{t-1}$ will then approach any particular one of these only if the series of $Y$'s is initiated by certain values. It is also possible that for certain special initial conditions $Y_t/Y_{t-1}$ may approach a value $r_i$ different from any of the dominant $r$'s, but that situation will be unstable. Growth at the minor rate of growth is an example of this.
9. The sum of the investment period plus one income period is called the income-investment period.

*accelerator (B) multiplied by the number of income periods per year (m).*[1]

That is, the dominant component will have a growth factor at least equal to $mB$ per income-investment period. If then the ultimate accelerator, $B$, is of the magnitude of 1 to 2, and the income period about $\frac{1}{3}$ year ($m = 3$), then dominant steady growth must be of the magnitude of 200 to 500% per income-investment period. Unless $B$ is very small, or the income-investment period very long, steady growth will be associated with a very large dominant rate.

2. *Values of the accelerator and marginal propensity to consume and their periods which will yield steady growth are not entirely implausible even though they lead to tremendous dominant rates of growth.*

CRITICAL COMBINATIONS OF $a$, $c$, AND $n$ FOR STEADY GROWTH



C, CURRENT ACCELERATOR
c = b ÷ investment period in years
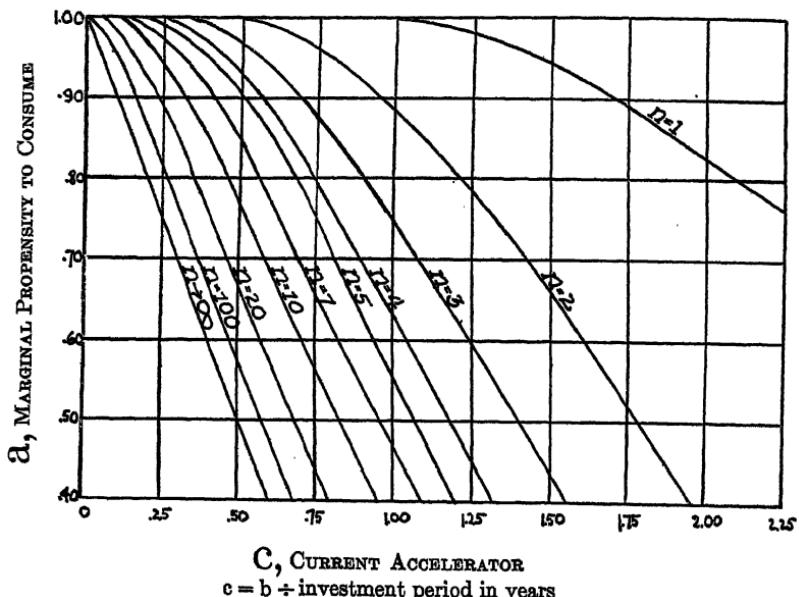
For any specified set of $a$, $c$, and $n$, conditions I and II tell us whether steady growth is possible. The Chart shows the minimal combinations of these three quantities that will permit steady growth. Each curve on the chart shows for a given $n$ the smallest value of $c$ that will suffice to yield steady growth for any given value of $a$, or the smallest value of $a$ that will yield steady growth for any given value of $c$.

All combinations of $a$ and $c$ that lie above or to the right of the

1. This follows from condition IV since, by definition, $nc = mB$.

curve for a given $n$ will lead to steady growth for that value or any larger value of $n$. Thus for $n = 10$, and $a = .85$ steady growth will be achieved if $c \geqslant .40$. These minimal values of $n$ and $c$ might, for example, correspond to an ultimate accelerator of 1.33 operating over $3\frac{1}{3}$ years, and an income period of $\frac{1}{3}$ year. Or, possibly, they might correspond to an ultimate accelerator of 2.0 operating over 5 years and an income period of $\frac{1}{2}$ year. For these examples the dominant rate of steady growth will exceed 300% per income-plus-investment period.

Or suppose the marginal propensity to consume = .80, the income period $\frac{1}{3}$ year, and the ultimate accelerator 1.55 operating over $3\frac{1}{3}$ years. Then steady growth will dominate. But in this case the rate of growth of the dominant component would be over 400% per $3\frac{2}{3}$ years. Or if $a = .50$, the income period $\frac{1}{3}$ year, and the ultimate accelerator 2.8 operating over $3\frac{1}{3}$ years, there will be steady growth. The reader may judge for himself whether these or other possible combinations which lead to steady growth are implausible.

3. *Only if the marginal propensity to consume is very close to unity can dominant steady growth be moderate.*

This may be seen from the chart, for the dominant growth factor per income-investment period is at least $nc$, according to condition IV. We can hardly expect the investment period to be as short as the income period, so that $n$ must be at least equal to 1, and is probably considerably greater than 1. Nor should we expect the investment period plus the income period to be over 6 years. If the dominant component of income is to grow steadily at a rate of less than 50% increase per 6 years, then the marginal propensity to consume must be greater than .95. If the investment period were shorter, so that, let us say, the income-plus-investment period is about $2\frac{1}{2}$ years, then if the dominant component is to grow at less than 20% per $2\frac{1}{2}$ years, $a$ must be greater than .98.

For values of $a$ close to .80, the dominant rate of steady growth, if steady growth is at all possible, must be over 106% per income-investment period. For values of $a$ not close to unity, moderate minor rates of steady growth are possible, but these will be paired with tremendous dominant rates of growth. Accelerator and propensity may permit steady growth, but eventually only at very high rates of growth of the dominant component; moderate initial rates of growth will soon be overshadowed by the dominant component unless stabilized by factors not yet considered.

QUALIFICATIONS

### 1. *Non-linearity of accelerator and propensity.*

If, as the rate of growth of income increases, the values of the accelerator and propensity change, might not a moderate dominant rate of steady growth be achieved? This requires that $a$ approach unity and that $nc$ remain close to unity. Such behavior is conceivable, but unlikely. It is much more probable that, as income grows, the marginal propensity to consume will remain well below unity; and $c$ may even be expected to increase, as is argued below. It is not then probable that as income grows the values of $a$ and $c$ will change so as to moderate the dominant rate of growth.

### 2. *Additional expenditures.*

There may well be other expenditures in addition to those taken account of in equation (5). For example, there may be some consumption independent of the level of national income, or some investment either independent of national income or dependent on its level rather than on its rate of change. All such additional expenditure can be broken down into four parts:

1. One part linearly dependent on the national income of the previous period (or periods[2]).
2. One part linearly dependent on changes in the national income over the past.
3. One part constant.
4. One part variable over time, and linearly independent of both the income of the previous periods and changes in incomes of previous periods.

Expenditure other than consumption which is linearly dependent on the national income can be lumped with consumption so that $a$ may be redefined, not as the propensity to consume, but as the propensity to spend. Similarly any expenditure linearly dependent on the change of the national income can be lumped with investment. The only characteristic of investment that we have relied on in this paper is that it is linearly dependent on changes in the level of national income. This is probably true of only part of investment so that what we have been calling investment might better be termed accelerator expenditure, which should be defined to include all expenditure linearly dependent on the change of income.

If expenditure which is constant over time, and therefore inde-

---

2. Dependence of expenditure on income of several preceding periods is considered in the discussion of the general linear case in the appendix.

pendent of the rate of income, amounts to $b$ per time unit, then it will be multiplied and contribute $\dfrac{b}{1-a}$ to the national income of any period. It is interesting to note that the multiplier of constant expenditure depends only on $a$, the propensity to spend, and not at all upon the accelerator. ($a$ is assumed less than 1.) We may then redefine $Y_t$ to be the national income minus $\dfrac{b}{1-a}$, its constant component.

These three sorts of expenditure can accordingly be included in an equation of the form of (5) without requiring any change except a redefinition of terms. This is not true of the fourth type of expenditure. Let us call the fourth type of expenditure, that which varies independently over time, $z_t$. Introducing this into (5) we get:

$$Y_t = a\,Y_{t-1} + c(Y_{t-1} - Y_{n-1}) + z_t \qquad (6)$$

The new $Y_t$ will have the same components as one determined by (5) plus a new component introduced by the $z$'s.[3] The national income then becomes a weighted sum of all the $z$'s up to date, whose weights depend on $a$ and $c$. If, in particular, $z_t$ has a regular time shape, such as a sine curve, or a steady rate of growth, then $Y_t$ will include a component of similar time shape multiplied by a certain factor. For example, suppose that $z_t = kg^t$. It will contribute a component $\lambda g^t$ to $Y_t$ where $\lambda = \dfrac{k}{g^{n+1} - (a+c)g^n + c}$. If $g > 1$, then $z_t$, the independent steady growth of expenditure, will contribute a component of steady growth to $Y_t$. In this case the multiplier, $\lambda$, does depend on the accelerator as well as on the propensity to spend. If the independent rate of growth, $g$, is close to one of the growth factors of (5), then the multiplier $\lambda$ will be very large.

Thus even though the values of accelerator and propensity are not such as to generate steady growth, they may lead to a large multiplier operating on steadily growing expenditure which results from such independent factors as increase of the population, changes of taste, new inventions, etc.

3. For a discussion of the nature of the solution when $z$ is a random or stochastic variable see: Herman Wold, *A Study in the Analysis of Stationary Time Series* (Uppsala, 1938), or Paul Samuelson, *Foundations of Economic Analysis* (Cambridge, Mass., 1947). Both these discussions are based on the assumption that the difference equation, excluding the $z$'s, would not yield steady growth.

### 3. *Expectations.*

It may be argued that investment depends upon expectations of future income rather than upon past changes in income. The inclusion of expectations into our system introduces no further complications. For any expenditure related to expectations can be included in the category of "additional expenditures" and can be broken down into the four categories of additional expenditure listed above.

Consequently the operation of expectations may change the values of the constants in (6) or change the time-shape of the $z$'s, but no further change is introduced.

We must accordingly interpret $a$ as including all expenditures proportional to the level of income, including the expenditures brought about by the influence of the level of income on expectations. Similarly $c$ includes the effects of changes in the level of income on expenditures, including the expenditures brought about by the influence of changes in the level of income on expectations. Finally $z_t$ must include expenditures which depend on changes of expectations which are independent of the level of national income and changes thereof.

### PRICE AS A STABILIZER

The steady growth that is indicated by equation (5) when conditions I and II hold is a steady growth of expenditure. The expenditure is assumed to equal income at current prices. Should increased expenditure lead to price rises, equation (5) can no longer be expected to hold at current prices. An increase of money income when prices are rising cannot be expected to lead to the same investment as an equal increase of money income when prices are constant. On the other hand, equation (5) need not hold uniformly in real terms irrespective of what is happening to prices, since price changes may alter the consumption and investment patterns.

These considerations are especially important when full employment is approached, and only a very moderate rate of steady growth of real income is physically possible. When unemployment exists an increase of production can be achieved by hiring unemployed workmen and putting them to work. The capital requirements may in part be met by existing capital so that new investment need not equal the amount of capital used by the re-employed workers. As production increases, however, further re-employment requires larger investment; as there is less and less idle equipment, the needs for new investment are greater. Then as full employment is approached

increases of production in excess of those associated with the natural growth of the labor force can be achieved principally by the transfer of workers from tasks of low productivity to tasks of high productivity, and that usually requires large additions to capital. Finally the productivity of labor in any particular task can usually be increased by the use of more elaborate machinery. Obviously much more investment will be required to support a given increase in production when capital alone is increasing than when employment and capital are increasing together. Consequently the value of $c$ will probably be very large at full employment.

With a large $c$, the dominant rate of growth of expenditure indicated by equation (5) would be very great. But physical limitations would prevent a rate of growth of real income by more than a few per cent a year. What then will happen? A moderate price rise may either preserve the conditions of moderate steady growth or lead to cyclical fluctuations.

Let us, for the moment, assume that price changes do not affect either real consumption expenditure or real investment expenditure; and that the latter depend only on the level of real income or on changes in that level. Let us also abstract from independent expenditure such as $z_t$ in equation (6). Then, if we denote real income of the $t$'th period by $\overline{Y}_t$, and the price level by $p_t$, so that $\overline{Y}_t p_t \equiv Y_t$, the real expenditure of the $(t-1)$th period will be $a\overline{Y}_{t-1} + c(\overline{Y}_{t-1} - \overline{Y}_{t-n-1})$. This becomes income in the $t$'th period, since we have taken the income period as our time unit. Then the money income of the $t$'th period will equal the real expenditure of the $(t-1)$th period valued at the price level of that period, so that

$$Y_t = p_{t-1}[(a+c)\overline{Y}_{t-1} - c\overline{Y}_{t-n-1}] \tag{5a}$$

or

$$\overline{Y}_t p_t = p_{t-1}[(a+c)\overline{Y}_{t-1} - c\overline{Y}_{t-n-1}] \tag{5b}$$

or

$$\overline{Y}_t \left(\frac{p_t}{p_{t-1}}\right) = (a+c)\overline{Y}_{t-1} - c\overline{Y}_{t-n-1} \tag{5c}$$

It may be observed that equation (5c) is similar to equation (5) except that $\overline{Y}_t$ is substituted for $Y_t$ throughout, and that there is a relative price factor on the left hand side. Therefore whatever rate of growth is indicated by the values of $a$, $c$, and $n$ will now be broken down into increase of real income and increase of price level. The rate of growth of real income will be reduced not only by the immediate rise in price, but also by the smaller investment that will result from the fact that real income does not go up as fast as money income.

As an example, let us insert in (5c) the same constants as we considered in (2a) above. In that case they yielded minor and dominant rates of growth of 5% and 100% respectively.

$$\overline{Y}_t \left( \frac{p_t}{p_{t-1}} \right) = .95 \, \overline{Y}_{t-1} + 2.1(\overline{Y}_{t-1} - \overline{Y}_{t-2}) \qquad (5d)$$

Let us assume that a 6.9% increase in money expenditure leads to a 6% increase in production and to a 0.83% increase in price level. Then equation (5d) will be satisfied by a 6% annual rise in real income and an 0.83% annual rise in prices, as may be seen by substituting $\overline{Y}_t = K(1.06)^t$ and $\left( \frac{p_t}{p_{t-1}} \right) = 1.0083$ in (5d) where $K$ is any constant.

This rate of growth of 6% will be a stable rate of growth[4] provided that larger rates of growth lead to sufficiently larger price rises so that the higher rates cannot be maintained, and that smaller rates of growth lead to smaller price rises such that the rate of growth will be increased.[5] Consequently a very gentle price rise can stabilize at or about 6% a year a rate of growth that would with constant prices approach 100% a year. But if price rises are slightly more vigorous, then they may cause the time shape of income to switch over from steady growth to cycles. Thus if in (5d) any increase in money expenditure would be split equally between price rise and increase in real income, then (5d) would indicate a cyclical fluctuation.

Since $Y_t$ is not measured from zero level of income, but from an equilibrium level, it is unrealistic to assume that a given percentage change in $Y_t$ will lead to a given percentage change in prices. For values of $Y_t$ close to zero a 5% or 10% change in $Y_t$ might be negligible and so lead to no significant price change, while for $Y_t = 100$ billions, a 5% change might lead to a large price change.

---

4. In this connection by a rate of growth stable within a given region is meant that if we have as initial conditions a given rate of growth within the specified region subsequent rates of growth will gradually approach the stable rate of growth.

5. In equation (5d) if $\dfrac{p_t}{p_{t-1}} > \dfrac{Y_t}{Y_{t-1}} \left( \dfrac{3.05 - \dfrac{Y_t}{Y_{t-1}}}{2.1} \right)$ for $\dfrac{Y_t}{Y_{t-1}} > 1.069$, and if the inequality sign is reversed for $\dfrac{Y_t}{Y_{t-1}} < 1.069$, then 1.069 is a stable growth factor of money income. This inequality implies, for example, that a 10% increase in $Y_t$ induces more than a 2.1% increase in prices, while a 6% increase in $Y_t$ induces less than a 0.45% increase in prices.

Consequently, with a given $a$, $c$, and $n$, $Y_t$ may have a steady rate of growth over a certain range; then as further changes in $Y_t$ bring price rises, the rate of growth may be moderated; and finally $Y_t$ may reach a range in which cyclical fluctuations are generated.

If over any portion of this range of fluctuation increases of price have the effect of stimulating speculative investment in inventories or other assets, this tends to increase the accelerator so as to modify or overcome the moderating effect of price changes.

The minor rate of growth, as we have seen, constitutes the critical dividing line between rates of growth that lead to expansion and those that lead to decline. If the full-employment limit on the increase of production is greater than the minor rate of equation (5), then a steady rate of growth of real income at the full employment limit can take place, accompanied by a steady, but possibly moderate rise in prices. Thus in equation (4) above, the minor rate of growth is slightly over 3% per year. So equation (4), which would lead to a rate of growth of over 150% a year if prices did not rise, will be satisfied by a production rise of about $3\frac{1}{2}\%$ a year accompanied by a price rise of 1.5% a year.

In conditions of full employment the minor rate of growth becomes the critical quantity, and the dominant rate may be relegated to the background. The lower the minor rate the better the chances that a steady rate of growth may be maintained at full employment. Those values of $a$, $c$, and $n$, which will yield a large dominant rate of growth will also yield a small minor rate of growth. In particular a high value of $c$ will lead to a small minor rate. Consequently, if, in fact, $c$ is very large at full employment, steady growth may be possible at the moderate rate permitted by physical limitations.

We may review briefly the manner in which a price rise acts as a moderator: increasing real income stimulates investment expenditure and consumption expenditure which together increase aggregate monetary expenditures. If increased monetary expenditure can bring out only a certain limited increase in output, the rest of the increase in expenditure will merely raise prices. As a consequence real income will increase only moderately and so there will be induced only a moderate rate of real investment, which will then bring out only moderate increases in output and in prices. If the physical limit to the rate of growth of real income is not much greater than the minor rate of growth determined by $a$, $c$, and $n$, then the price rise need only be small. If the physical limit is much above the minor rate of growth, then the price rise will be somewhat larger. Thus in

equation (5d) if the physical limit of the rate of growth of real income were as high as 10% per income period instead of 6%, then the price rise indicated would be 3.7% a year instead of .83%. In general, for any reasonable upper limit of growth of real income at full employment, only a gentle price rise will be required as moderator.

We have at last achieved our goal of finding the conditions that will yield moderate steady growth. The first requirement is that steady growth be possible at all, which means that propensity and accelerator must be great enough, or income or investment period short enough, so that $a$, $c$, and $n$ satisfy conditions I and II. Furthermore $a$ or $c$ (or both) must be sufficiently greater than required by conditions I and II (or the income period sufficiently shorter) so that a moderate *minor* rate of steady growth is possible. This will imply a tremendously high dominant rate of steady growth unless $a$ is close to unity. Then, if not stabilized by price, any rate of steady growth above the minor rate will lead to accelerated growth approaching the dominant rate.

The second requirement is, therefore, that increases of expenditure lead to some increase of prices as well as to increases of production. Then the tendency to approach the dominant rate may be stabilized at a rate of growth quite close to the minor rate.

The stabilizing effect of rising prices is not limited to relationships of the type of equation (5). Consider any set of conditions such that, if real income has increased over the relevant past[6] with a growth factor $r$, expenditure in the current period will exceed that of the immediately preceding period by a factor greater than $r$. It is obvious that there exists a price rise which will just reduce the real value of the increased money expenditure to $r$ times the real income of the previous period. Such a price rise would then bring an end to further increase in the rate of growth. If prices rise when expenditure grows at a rate close to the minor rate, then the equilibrium rate of growth of real income may be very close to the minor rate. For we have seen that, for initial rates of growth very close to the minor rate, the immediately following rate of growth will exceed the initial rate by a small amount, and so only a very small price rise will be required to keep the growth of real income down to the initial rate. A larger price rise may eliminate completely the possibility of steady growth.

This result is quite general, so that in almost any case in which a moderate rate of growth is at all possible, the instability introduced

6. The relevant past includes those preceding periods whose incomes affect expenditure in the current period.

by the dominance of much greater rates of growth may be overcome and the situation stabilized by a moderate price rise, so long as the price rise itself does not stimulate additional real consumption or investment.

SIDNEY S. ALEXANDER.

HARVARD UNIVERSITY

## APPENDIX

*The Conditions for Steady Growth.*

Given (5) we may write down its characteristic equation:[7]

$$f(x) \equiv x^{n+1} - (a + c)x^n + c = 0. \tag{7}$$

The dominant component of (5) will grow steadily if the root of (7) with largest absolute value is real and greater than unity. By Descartes' rule of signs, (7) may have either two or no positive real roots. Note that $f(x) > 0$ for $x = 1$,[8] and for $x = a + c$; and that at the latter value all the derivatives of $f(x)$ are positive. Therefore no real root is as great as $a + c$, and there are either two or no real roots between 1 and $a + c$. If there are two such roots in that interval, then over the interval $f(x)$ must decline to some minimum value less than zero and then again increase to the positive value $c$ at $x = a + c$.

But $f(x)$ has a minimum where $f'(x) = 0$ or $x_{\min} = \dfrac{n}{n + 1}(a + c)$

7. This equation is derived by substituting $x^{n+1-i}$ for $Y_{t-i}$ in equation (5). It is useful because the general solution of (5) is $Y_t = \sum\limits_{i=1}^{n+1} K_i r_i$ where the $K$'s are constants depending on the initial sequence of incomes that starts the series off, and the $r$'s are the $n + 1$ roots of the characteristic equation. (The $K$'s might be periodics of period 1.)

If an $r_i$ is real:

$r_i > 1$ contributes a component that grows steadily;

$0 < r_i < 1$ contributes a component that steadily approaches zero;

$-1 < r_i < 0$ contributes a component that approaches zero with damped oscillations of period 2;

$r_i < -1$ contributes a component that steadily grows in absolute value and oscillates each period from positive to negative or negative to positive.

If an $r_i$ is complex: It contributes a component that oscillates in either a damped or explosive sine curve; damped if $|r_i| < 1$, explosive if $|r_i| > 1$.

8. Since $f(1) = 1 - a$, and by assumption $a < 1$. If $a > 1$ then $f(1) < 0$ and there must then be a root of $f(x)$ between 1 and $a + c$. This proves that a marginal propensity to consume greater than unity always leads to a component with steady growth.

This minimum is in the interval

$$1 < x < a + c \quad \text{if and only if,} \ \frac{n}{n+1}(a+c) > 1. \qquad \text{(I)}$$

Substituting $x_{\min}$ in $f(x)$, the condition that at this minimum $f(x) \leqslant 0$ gives:

$$nc \leqslant \left[ \frac{n}{n+1}(a+c) \right]^{n+1} \qquad \text{(II)}$$

The larger real positive root, $r$, will lie between $x_{\min}$ and $a + c$,[9] so that

$$\frac{n}{n+1}(a+c) \leqslant r < a + c \qquad \text{(III)}$$

But II and III together imply:

$$nc \leqslant r^{n+1} \qquad \text{(IV)}$$

By definition: $c = B$ divided by the investment period in years; $n$ = ratio of investment period to the income period. Then if there are $m$ income periods per year, the investment period is $\frac{n}{m}$ years.

Then

$$c = \frac{B}{n/m} = \frac{mB}{n} \quad \text{and} \quad nc = mB$$

so that, by IV:

$$mB \leqslant r^{n+1} \qquad \qquad \cdot$$

So far we have established merely the conditions that there be a real positive root greater than unity. We may now show that this root is the root of greatest absolute value.

Divide (7) by $(x - r)$ where $r$ is the larger positive root, but notice that $(a + c - r) = \dfrac{c}{r^n}$ since $r$ is a root of $f(x)$:

$$\phi(x) \equiv \frac{f(x)}{x-r} = x^n - \frac{c}{r^n}x^{n-1} - \frac{c}{r^{n-1}}x^{n-2} \cdots - \frac{c}{r} = 0 \qquad \text{(8)}$$

But (8) is the characteristic equation of a difference equation:

$$Z(t) = \frac{c}{r^n}Z(t-1) + \frac{c}{r^{n-1}}Z(t-2) + \ldots + \frac{c}{r}Z(t-n) \qquad \text{(9)}$$

---

9. More accurately the larger root must lie either at $x_{\min}$ or between $x_{\min}$ and $a + c$. If the larger root is at $x_{\min}$ it then coincides with the smaller root, and there is a multiple root at $x_{\min}$.

All of the coefficients on the right are positive. If we start off with a set of $n$ $Z$'s, each $\geqslant 0$ and not all equal to zero, then all subsequent $Z$'s will remain $> 0$. It is possible to choose an initial set of non-negative $Z$'s such that the coefficient of the largest root in the solution of (9), $Z_t = \sum_{1}^{n} K_i r_i^t$, is not zero.[1] If the dominant root were negative or complex, then any solution of (9) in which the coefficient of the dominant root was not zero would lead to some negative values of $Z_t$, even if the initial $Z$'s were positive. Since $Z_t$ always remains positive so long as the initial $Z$'s are non-negative and not all equal to zero, the dominant root (or one of the dominant roots if there are several of equal absolute value) must also be positive. But since $f(x)$ has only two positive roots, $\phi(x)$ has only one, so we see that the minor positive root dominates all the remaining roots of $f(x)$, or is at least as great in absolute value as the largest remaining root. The larger positive root is accordingly the dominant root of $f(x)$.

*Neglect of the Income Period.*

Consider equation (5) with given investment period in years as the income period approaches zero. Then $n \rightarrow \infty$, and if $(a + c) > 1$, conditions I and II will be satisfied for $n$ greater than a certain value. Then condition IV indicates that $r^{n+1} \rightarrow \infty$, so that the dominant growth factor for the income-investment period becomes infinitely large as the income period approaches zero. But if we ignore the income period completely:

$$Y_t = (a + c)Y_t - cY_{t-n'} \tag{10a}$$

where $n'$ is the investment period measured now in arbitrary time units, say months. The characteristic equation is:

$$x^{n'}(1 - a - c) + c = 0 \tag{10b}$$

whose solution is obviously

$$x = \left(\frac{c}{a + c - 1}\right)^{\frac{1}{n'}}$$

If $a + c > 1$ and $a < 1$, the absolute value of each of the $n'$ roots will be greater than 1, but will decline toward unity as $a + c$ increases, a form of behavior characteristic of the minor roots. Equation (10b) has one less root than equation (7). The root that has been lost is

1. In particular we may choose $Z(t - 1) = 1$ and $Z(t - i) = 0$ for $i > 1$ and no $K_i$ will vanish so long as there are no multiple roots.

the dominant root of equation (7) which has become infinite; thus neglect of the income period loses the dominant root from sight.

*The General Linear Case.*

The generality of equation (5) is limited by the assumption that there is a single income-period and that investment operates evenly over the investment period. The general linear case of determination of the level of income by expenditure dependent on previous income, and by investment dependent on changes in income is:

$$Y_t = \sum_{i=1}^{p} a_i\, Y_{t-i} + \sum_{i=1}^{n} c_i(Y_{t-i} - Y_{t-i-1}) \tag{11}$$

where the time-unit is some specified period — say one month. The conditions required for a dominant component yielding steady growth are quite complicated in this case, but certain general conclusions carry over from equation (5). The general principles that if steady growth is possible the dominant rate will be very large unless $a$ is very close to unity, and that moderate steady growth can then be stabilized by a moderate steady price rise, still hold.

In this connection $a$ must be defined as the ultimate marginal propensity to consume, or the sum of the $a_i$'s in (11). Also, if (11) yields steady growth the dominant rate will be reduced for a given $a$ if the later expenditures are increased at the expense of the earlier ones. This corresponds to a lengthening of the average income period, or a reduction of the income-velocity. But as in the earlier cases, for $a$ not close to unity it is not possible by lengthening the average income period to reduce the dominant rate of steady growth down to a moderate rate, but before that happens (11) will yield a cyclical dominant component rather than one of steady growth.

The relation of (11) to (5) can perhaps best be shown by defining a weighted marginal propensity to consume, and a weighted average current accelerator for given growth factor $r$ as:

$$\bar{a} = \sum_{i=1}^{p} \frac{a_i}{r^{i-1}} \qquad \bar{c} = \sum_{i=1}^{n} c_i\, r^{n-i} \Big/ \sum_{i=1}^{n} r^{n-i}$$

If $r$ is the dominant root of the characteristic equation of (11) the equation:

$$Y_t = \bar{a}\, Y_{t-1} + \bar{c}\, (Y_{t-1} - Y_{t-n-1}) \tag{11a}$$

will have the same dominant component as equation (11). If that dominant component is to yield steady growth, then $\bar{a}$ and $\bar{c}$ and $n$ must fulfill conditions (I) and (II). Thus many of the conclusions

of the earlier cases carry over. In particular unless $\bar{a}$ is very close to unity, steady growth will be possible only at a very rapid rate. In any case the growth factor, if steady growth is possible, must be greater than $n\bar{c}$ per period of $(n + 1)$ time-units.

Or we might have considered an ultimate marginal propensity to consume and average current accelerator independent of a particular weight:

$$a = \sum_{i=1}^{p} a_i \qquad \text{and} \qquad c = \frac{1}{n} \sum_{i=1}^{n} c_i$$

Then we can redefine the time unit and the investment period $(\bar{n})$ to be a weighted average of the lag of consumption behind income and of investment behind changes of income respectively. The weights depend in a particular way on $r$, the dominant root of the characteristic equation of (11), so that the relationship:

$$Y_t = a\,Y_{t-1} + c\,(Y_{t-1} - Y_{t-\bar{n}-1})$$

will have the same dominant component as does (11). Therefore, if (11) is to have a principal component which grows steadily, $a$, $c$, and $\bar{n}$ must satisfy conditions (I) and (II). The principal conclusions are accordingly valid for the most general linear case. Since non-linear cases can, of course, be approximated over limited ranges of variation by linear relationships, our conclusions are quite general.

# CENTRAL BANKING IN THE LIGHT
# OF RECENT BRITISH AND AMERICAN EXPERIENCE

## SUMMARY

I. Introduction, 198. — II. Central bank control during the inter-war period, 198 — III. More recent problems: control of long-term interest rates in Britain and the United States, 201. — IV. Selective credit policy as a supplement to low and stable interest rates, 206. — V. This policy not a complete break with the past, 209. — VI.

## I

In time of war, central banking, like so many other things, is pushed aside or at least loses all its virtue. Under the impulse of war finance, the technique of the central bank is manipulated so as to conform to events rather than used to control them. Central bankers become the slaves of deficit finance. But the war has now been over for three years and government deficits have more or less disappeared, both in the United States and the United Kingdom, and in many other countries as well. Central banking has at last reared its head once more, though in a world much changed by the fashion of full employment policies and the wide extension of government economic activity. In the new atmosphere there is inevitably an inclination to deprecate the usefulness of traditional central banking. But the progress of post-war inflation has led to some questioning both in the United States and in Britain: has the reaction against the traditional financial weapons for fighting inflation perhaps gone too far? Is there after all some useful part for the central bankers to play?

A reconsideration of the fundamental nature of central banking therefore seems timely. I do not attempt in this paper to answer all the contemporary questions about the control of inflation. What I am going to do is to review the development of central banking during the last two decades, drawing attention to the swing away from the traditional "quantitative controls" and towards "selective credit policies"; to put this change of emphasis against the wider setting of the earlier history of central banking; and to consider what, in the light of all this, can be said about the future.

## II

The inter-war period can perhaps be called the heyday of central banking. It was certainly the period in which central banking was

most widely talked about; and (since the theory is important) it is worth recalling the terms in which authorities did then talk about the functions of central banks. It is unnecessary to burden the reader with numerous quotations: I think I can portray the position briefly in very general terms. The Macmillan Committee spoke of a central bank as being "called upon to keep the financial structure upon an even keel," and went on to explain that it must have power to control the expansion and contraction of credit, and to emphasize the paramount importance of open market operations and rediscounting at interest rates fixed by the central bank. This authoritative analysis fairly represented informed opinion in the United Kingdom, and I venture to say in the United States as well. That is to say, we thought in those days of bank rate and open market operations — the control of the quantity and of the price of credit — as the basic work of a central bank.

Before discussing a certain swing of opinion away from the confidence and simplicity of this earlier view, I want to take note of the ways in which the last twenty years have seen the full development — I think I may say the perfecting — of the technique of control over the quantity and price of credit, at any rate in the United States and in the United Kingdom.

In the United States, the banking crisis of 1933 revealed great weaknesses in the system: notably (for our present purpose), the power of the Federal Reserve System to pump cash into the commercial banks did not prove equal to the task. In legislation since then (particularly the post-crisis legislation of 1933 and 1935), the Federal Reserve System has been empowered to take up and hold much more freely government paper; and a second war has called into existence a greatly increased volume of such paper. The Federal Reserve System has also been empowered to enlarge indefinitely (if need be) the previously narrow limits of paper eligible as collateral for advances by Reserve Banks to member banks. The constitution of the System itself has been changed, power being now more clearly concentrated in the central Board of Governors, so that there is now more chance of the quick decisive action which a crisis necessitates. A most important innovation is the grant of power to alter the cash reserve ratios which are the link between Reserve Bank cash and the total supply of commercial bank deposits. The theorists of the twenties had seen in these cash ratios a foundation of central bank control of member banks: the legislators of the thirties greatly increased central bank powers by the grant of this new weapon,

whereby the central bank might by a stroke of the pen virtually halve or double the monetary base.

The stark reality of this new power did frighten some people, and the suspicion that it was too blunt and heavy an instrument for the delicate business of financial control was perhaps heightened by the early use of it in efforts to deal with the excess reserves problem of 1936–38.  This criticism is warranted only against sweeping changes in the legal cash ratios — raising by large and round percentages such as 50 or 100.  But there is no reason why the percentage changes should not be in nicely-calculated fractions, which would be precisely equivalent in their quantitive effect to open-market operations which no one would think inappropriate.  Indeed, the alteration of reserve ratios has some advantage over the open-market operation, in that the impact effect on particular banks can be precisely foreseen, whereas the impact of open-market operations is a matter of guess-work, especially in a unit banking system of the American type.

The Governors of the Federal Reserve System are not satisfied by the great extension of their power over the total supply of money, and among the new powers they have been persistently seeking from Congress are two directly concerned with this traditional aspect of a central bank's main work.  They have been seeking a freer hand in the alteration of cash reserve ratios, and the power to prescribe minimum ratios of short-term government paper to be held by the commercial banks.  Although the latter proposal found its origin in the particular shape of post-war bank statistics, it is fundamentally to be ascribed to the desire to secure complete control over the quantity and price of credit.

In the United Kingdom also, there has been, over a rather longer period, a perfecting of the power of the central bank to control the quantity and price of credit.  The major advance of the century resulted from the enlargement, in the first war, of the supply of Treasury Bills — the ideal security for open market operations.  This by-product of government war finance provided a solution for the pre-1914 problem of market control; and from 1919 onwards the Bank of England was always able to make itself felt rapidly and easily.  During the thirties the authorities, coping with the large movements of foreign funds, became a great deal more agile in their open market operations, and this agility, combined with the prestige of the Bank of England under Montagu Norman, had produced an atmosphere in which ready co-operation of the market was the natural corollary of acknowledgement of the Bank's leadership.  In these circumstances it became unnecessary for the Bank ever to "force

the market into the Bank" — i.e., to make the official bank rate an effective dealing rate, and we were soon able to drop out of the habit of learning at Thursday lunch-time what had happened to bank rate. The intervention of the "special buyer" in the market ceased to be a subject for extraordinary comment: the contact between the Bank of England's operator and the bill-brokers became virtually an everyday affair. Indeed, before the end of the thirties direct discounting contacts between the joint-stock banks and the Bank of England had reappeared — in breach of an eighty-year-old tradition.

Despite this advance to a smooth and confident technique of Bank of England control, there still remained in the English system a certain element of play which was a potential source of embarrassment for the authorities. This was the slight flexibility of the commercial banks' cash ratios, which were fixed only by convention. They had been known (in the latter twenties) to vary significantly, quite openly; still more, they could and did vary substantially behind the facade of window-dressing. The Macmillan Committee had pilloried this practice as far back as 1931, but it went gaily on for another fifteen years. Since 1945 the authorities have found ways of dealing with this element of slack in the link between the central bank and the commercial banks. In the first place, central control was increased by the retention, after the disappearance of war conditions, of the Treasury Deposit Receipt. As long as the T.D.R. could be used as cash in subscribing to new government bond issues, it was virtually the same as cash to the commercial bank; but when wartime bond issues ceased, T.D.R.'s lost their original liquidity and (since their amount is dictated by the authorities) became yet another means for absorbing or releasing cash to the reserves of the commercial banks. More fundamentally, the authorities took early advantage of the latent power of the Bank of England Act of 1946 to stop window-dressing and virtually fix the cash ratio at eight per cent. The total effect of these changes of recent years is thus to tidy up and tighten the control of the Bank of England over the quantity and price of bank credit, a control already substantially achieved by the technical developments of previous decades.

## III

The latest decades have thus seen a consolidation of the power of the world's greatest central banks — of their power, that is to say, to control their respective monetary systems. To what end have they directed their exercise of this power? It has certainly

not been the control of bank advances to business. Here I think we can see that there has been a tacit acknowledgement that the amount of ordinary bank advances to industry and trade is a relatively passive element in the economic conjuncture; and accordingly the central banks have always endeavored to allow the commercial banks to meet in full the "legitimate requirements" of industry and trade for temporary financial accommodation. In the face of the tendency of business to become less dependent on such accommodation (a tendency I have elsewhere called the secular decline in the demand for bank advances), and the central banks' willingness that these demands should all be met, the commercial banks have been in a constant position of wanting to expand their business loans beyond the limits apparently set by the demand: their "advances ratios" have been consistently lower than the banks would like to see them. The Keynesian "fringe of unsatisfied borrowers" has been absent.[1] The central banks have not put any pressure on the total of bank advances since 1929; and on that occasion attempts were made to confine the pressure to particular sections of the demand.

This attitude towards short-term bank advances accords with the dominant monetary theories: it marks the ascendency of the notion that variations in long-term capital investment, rather than in working capital, give the impulse to general economic fluctuations. While, therefore, the central banks have ceased to concern themselves with the amount of bank advances, they have directed their attention very actively to influencing long-term interest rates.[2]

In this pursuit of interest-rate policy there is a contrast to note between the experiences of the United States and Britain. In the United States the Federal Reserve System has exercised a more direct and much more successful control over long-term interest rates than has the Bank of England. The contrast in results derives partly from the fact that the Federal Reserve System does, whereas

1. This does not mean that there have been no unsatisfied borrowers, but that those unsatisfied have been below the standard required by the banks, a standard which would not be lowered in response to an increase in the cash basis. I understand the Keynesian fringe of unsatisfied borrowers to exist only if an increase in the cash basis would make the banks more willing to meet present demand for bank loans.

2. I should add that the Federal Reserve System has recently shown some perturbation at the difficulty it would encounter in forcing on the commercial banks greater caution in granting business loans: hence the desire for additional powers over the asset distribution of commercial banks. The System's concern over this shows that it has not abandoned all idea of direct quantitative control of bank advances.

the Bank of England does not, throw its weight directly and constantly into the long-term market as well as into the short. While the Bank of England effectively stabilizes the short-term rate by standing always ready to enter the market on either side, at the fixed interest rates, the Federal Reserve System has stabilized a whole "pattern of rates," short-, medium- and long-term, by standing ready to enter any section of the market in protection of the fixed rates. In both cases the relevant interest rates have been maintained by willingness to meet the public demand for money at the various rates of interest: the quantity of money has in this sense been fixed by the public (including the commercial banks) because the central banks have desired to fix the prices of money (the rates at which money exchanges against less liquid assets). In England, it is only the exchange relationship between money and short paper that has been fixed, because the central bank has dealt (consistently) only in short paper; whereas in the United States the central bank has been willing to vary also (to an amount sufficient to meet variations in outside demand) its holdings of longer government paper — and has therefore been able to fix the relationship between short and long government paper.

In this bolder activity the American authorities have run into some embarrassment because the outside valuation of long paper in terms of short has varied substantially. The "pattern of rates" chosen for freezing at the beginning of the second war was an abnormal pattern. It derived from the abnormalities of the post-crisis thirties, when a cheap-money policy was being pursued while long-term investment was checked by extreme uncertainties. In these conditions, short-term rates were below long-term rates to an altogether abnormal extent. Such an abnormal gap can persist only as long as people think that interest rates may rise. As long as the war emergency lasted, people could suppose the fixed pattern of rates to be a wartime phenomenon; and the suspicion that the outbreak of peace might be followed by higher rates sufficed to sustain a preference for short-term commitments — a preference measured by the abnormal margin between short and long rates. But when the war emergency passed, and the fixed pattern was still held, the supposition that it might be a merely wartime phenomenon gave place to the suspicion that the low rates on long-term bonds had come to stay. As soon as the market became confident that the Federal Reserve System would continue indefinitely to maintain the price of bonds (i.e., keep long rates down), long-term bonds became as good as shorts, indeed almost as good as money;

and the commercial banks (the principal outside operators in both shorts and longs) showed a disposition to unload shorts onto the Reserve Banks and to enlarge correspondingly their holdings of the more profitable longs. The market phenomenon evidencing the change of disposition was the tendency for prices of long paper to rise while the prices of shorts remained underpinned by Reserve Bank policy. A colossal mass of paper might have changed hands had the System attempted to hold the "pattern," and it had in effect either to acquiesce in the higher bond prices (lower *long* rates) or to raise *short* rates, in order to narrow the spread between them, in reflection of the outside rising confidence in the continuance of cheap money. The latter alternative — a slight rise in short rates — was chosen in 1947, and this, combined with a good deal of talk about the possibility of dearer money (shaking outside confidence in the level of bond prices, which the System would not underpin) sufficed to bring the pattern of rates and the market's preference for shorts against longs, once more into line with each other.

This kind of problem — sometimes referred to as the problem of monetization of the national debt — is one which has to be faced only by a monetary authority which ·is in a very strong technical position for operating directly on the long-term market. The Federal Reserve System has been in such a position, partly because its resources are large relatively to the total United States national debt. The Bank of England has not attempted to put itself in such a position, although, with the buttressing of the "Public Departments," it has sometimes come rather near to it. To say that "it has not attempted to put itself in such a position" perhaps implies that it could have done so had it tried. This is not quite the case: the relative size of the British national debt is so much greater, that the attempt by the Bank of England would, I suspect, have involved such a colossal actual (not merely potential) creation of money as dangerously to shake public confidence in the system.

This opinion is of course colored by the circumstances of the ill-starred British ultra-cheap money drive of 1946–47. In the nineteen-thirties the authorities had, by operating in the short market and by shrewd conversion operations, succeeded in driving the long-term rate down to the region of three per cent; and this rate was (with the aid of further propaganda, ample supplies of money, and physical controls) comfortably held throughout the war. Then in 1946–47 the authorities, by causing the rapid creation of bank deposits, by manipulating the security holdings of certain "Departments," and by propaganda in which the then Chancellor of the

Exchequer played a prominent part, forced the long-term rate down momentarily to about two and one-half per cent. The attempt to hold the rate at this low level failed fundamentally because the market was convinced that, in the light of the basic economic forces of high demand for capital investment and scarcity of resources, two and one-half per cent was an unnaturally low rate.

It is of course obvious that, in the prevailing economic circumstances, even three per cent was an unnaturally low rate, defended only by the stringent physical controls of capital investment. But three per cent had, after all, been held for years; and it was the attempt, the patently juggling attempt, to press it down yet further, flying in the face of basic conditions, that the market found so difficult to swallow. The distrust felt by the weighty investing institutions and by the operators (the two parties most influential in formulating the market demand schedule for government securities) was perhaps heightened by the tone of the propaganda speeches of the then Chancellor. At any rate, the demand curve for money (in exchange for fixed interest securities) proved both flat and unstable. The amount of new money pumped into the market (£900 million of bank deposits — a twenty per cent increase — in twelve months) was alarmingly high. The authorities then very properly got cold feet. Given market sentiment, creation of bank deposits on an enormous scale (an actual, as opposed to the American potential, monetization of the national debt) would have been necessary for holding the two and one-half per cent line; and the only alternative was retreat. The argument for retreat found support in the belief that the actual pressure of inflationary spending was being increased by the capital gains realized by sellers of long-term bonds, the taxed yield of which had become insignificant as compared with the untaxed capital gains.[3]

The authorities retreated and the yield on long-term government bonds slipped back to three per cent and even beyond. The scepticism of the market appears to have been fully justified, and this means that the authorities would have an even more difficult task on a future occasion. In 1946–47 the market could sometimes recall 1932 and think that perhaps the Chancellor would win after all; next time it will be all too easy to quote the ill-starred attempt of 1946–47. Thus the authorities, far from having proved that the power to create

3. The episode has been admirably discussed, with full factual analysis, by Mr. F. W. Paish in his article "Cheap Money" in *Economica*, August, 1947. He brings out particularly the way in which the lack of faith in the market, coupled with the technical conditions of taxed low dividends and untaxed capital gains, made the market extraordinarily unstable.

money *in unfrightening quantities* gives them power to enforce low and stable long-term rates of interest, have simply made the bond market peculiarly unstable — a happy hunting ground of speculators instead of the safe resort of "widows and orphans" and conservative investment institutions. Britain has had (and still has) cheap money; but it is a long way from having the "stable bond market" that has been successfully established by the Federal Reserve authorities.

The contrast between American success and British failure to achieve a fine degree of stability in the bond market has, I have suggested, been due basically to certain differences in technical conditions. The discredit into which the "ballyhoo" methods of an ex-Chancellor of the Exchequer have fallen is perhaps at the moment tending to exaggerate the limitations on the power of the monetary authorities; and perhaps thought should be given to the more technical causes of instability. After the experience of the last two years, greater stability in the bond market might be a more useful achievement than ultra-cheap money. No doubt the question is receiving attention in responsible quarters of the City of London. For what it is worth, I throw out the suggestion that something could be done by active co-operation between some of the great institutional buyers of gilt-edged securities — the banks, insurance companies, and investment trusts. These institutions have a strong interest in bond market stability, and their combined resources must be very large in relation to the turnover of the market. It is sometimes remarked that very few transactions are going through at times when prices are moving sharply; at such times concerted intervention (which would not necessarily have authoritative inspiration) by the big institutional investors could probably hold the line with ease. Having such notions as these in my head, I do not think the last word has been said about stabilization of long-term interest rates. In any event it is easy to exaggerate the British failure; it is only by contrast with the fine degree of control attained by the Federal Reserve System that the British case looks so hopeless. Taking a broad view, we can still say that both in Britain and in America traditional central banking has given way to the maintenance of conditions in which bond markets are held stable at low interest rates while there is always an adequate supply of banking funds available to meet the short-term credit requirements of business.

## IV

The stable bond market policy — the policy of maintaining interest rates at a steady low level — grew naturally in the thirties

out of the cheap money policy, the development reflecting a change of emphasis in contemporary monetary theory. When the slump of 1929–32 came, low interest rates were considered by theorists an important stimulus to the capital investment whose revival should pull the economy out of the slump. Its widespread adoption was hindered by the development of the liquidity crisis in 1931; but in Britain at least the cheap money policy secured governmental support because it implied a fall in the cost of servicing the national debt. Thereafter, very low interest rates were held for some years, but revival seemed painfully slow to materialize; and when it did come, it had a lop-sided appearance.[4] This experience of the limited efficacy of the cheap money policy no doubt helped to give force to the swing of theoretical opinion evident in the change of tone between Keynes's *Treatise on Money* (1930) and his *General Theory* (1936), and in the rise of the "stagnation thesis" in the United States. Direct enquiries into the reactions of business men to changes in the interest rate increased scepticism,[5] and faith in interest rate policy as a quick-working weapon declined. It became fashionable to stress the disturbing, rather than the curative, effects of rapid changes in interest rates, and to say that interest rates should be kept at a steady low level. Hence the acceptability of the "stable bond market policy" for peace as well as war.

If this line of thought were carried to its logical conclusion, central banks would become merely passive agents creating whatever cash basis was required to allow the continuance of the stable pattern of interest rates. As with the Issue Department as conceived by the Currency School of a century ago, the business of monetary control would be reduced to rule of thumb. But the central bankers, at least in the United States, have shown no disposition to retire thus easily from the fray; the decline in the importance of quantitative control has been paralleled by a rise (perhaps I should say, a revival) of the notion of qualitative control or *selective* control of credit. The basic argument for a selective credit policy is that while a low stable interest rate provides a favorable working background for capital development, the economy is still liable to run off the

4. Theoretically, interest rates should have their greatest effect on the demand for capital goods with the longest lives and to which the least uncertainty is attached. Dwelling houses meet these conditions while, especially when the industrial outlook is abnormally obscure, industrial equipment does not meet them. It is therefore not surprising that the British revival in the thirties centered largely around a housing boom, although there were of course other factors peculiarly favorable to such a development.

5. See for example, my article, "Business Men and the Terms of Borrowing," *Oxford Economic Papers*, May 1940.

rails more easily in some directions than in others, and that the central bank should use financial controls to check the unhealthy developments in these *particular directions.*

As a vague idea, selective credit control has a long history in every country where monetary thought has any history at all. But in the United States it has gained in weight from certain episodes, most of all perhaps from the boom of 1927–29. Two prominent features of that boom were the unhealthy extremity of the stock market speculation and the boom in durable consumption goods, the latter fostered by the support given by installment credit to the purchasers of the new durable consumption goods. The Wall Street boom was the more glaring of these two evils, and it had been a thorn in the side of the central bankers even while the boom lasted — owing to the prevalent notion that Wall Street was drawing credit away from its proper uses in legitimate industry and trade.

The Securities Exchange Act of 1934, because of these views, took a prominent place in the post-crisis legislation in the United States. It was easy, after the 1929 heart-burnings and the crisis heat, to pass legislation of this kind; but after more than a decade, the Federal Reserve System is still freely using the regulation of trading margins, and regards this as an important weapon in its armory. This is due to the more lasting swing of opinion away from a simple quantitative control of credit. The same may be said of the reiterated desire of the System for permanent powers, on the lines of the wartime emergency power, to regulate installment credit.[6] Governor Eccles and others have argued, with some force, that the variations in consumer credit are sizable in relation to total variations in national income, and that they finance a peculiarly unstable part of consumers' expenditure. There have also been suggestions of the desirability of regulating house-purchase finance, although central bank action in this field could hardly be necessary where government subsidies were important and could be varied as a part of a general employment policy.

These developments are indications of a swing in favor of selective credit policy as a supplement to the policy of low but stable interest rates. But if we are to have such regulations, why, it may be asked, should the work be put into the hands of the central banks? To be effective and equitable, any such regulatory functions must cover numerous corporations and individuals quite outside the banking systems of the country. This has been recognized in the United States both in the Securities Exchange Act and in the regulation of

6. In the Special Session of 1948, Congress revived the Federal Reserve System's power to control consumer credit.

installment credit. Here choice of the central bank as the regulating body may have been due to the convenience of using an existing body under whose peculiar constitution there is contact with business in every part of the country. But there is also, I think, a more logical basis for looking to the central bank for this work. The more I think about central banking, the more strongly I realize the cardinal importance of the work of diagnosis. Discretion is of the very essence of central banking, and if there is to be discretion there must be the utmost care in diagnosis. Now the diagnosis that should be the background of these new controls is diagnosis that the central bank should be making for its other operations also. Moreover, the knowledge that the central bank can gain from direct operation of these new controls is itself useful for enabling the central bank to build up the total picture of the economy which is essential for its general purposes. It seems to me entirely right, therefore, that these new controls, although they run far beyond the business of mere bankers, should be operated by the central banks. The fact that the central bank may itself be more or less directly controlled by the Treasury is irrelevant for this purpose; for the central bank is the body with the immediate contacts, with its finger all the time on the pulse of the body economic.

## V

What I have discussed in the last few paragraphs can perhaps be summarized in this way: The development of qualitative controls is in a sense the logical result of the adoption of a stable bond market policy. The old idea was to govern the pace of the economy by raising or lowering interest rates. Bitter experience has taught us that the interest rate, far from being the "delicate and beautiful instrument" it was once called, is a blunt instrument: it works too slowly in encouraging or discouraging in the right directions, and has its quick effects rather in the wrong directions. So we have moved towards the idea that the interest rate should be held pretty stable, and on the low side. And if this is done, it follows that the desired changes in activity must be encouraged, and the undesired changes discouraged, by other weapons. For instance, to discourage a particular boom we must tighten up on consumer credit, or on lending for the stock market, or whatever else is peculiarly appropriate to the particular conjuncture; meanwhile holding interest rates more or less unchanged.

That is, I believe, a fair enough short statement of the change of central banking ideas during the last generation. But now I want to go further back, and try to put this latest phase into its proper perspective in the light of the earlier development of central banking,

in the hope that to do so will stimulate central banking thought along fruitful lines.

The earlier history of central banking is of course almost entirely the history of central banking in England. On this subject much has been written: besides the distinguished official history by Clapham, there have been the valuable outside studies by Hawtrey, King, Morgan, Wood, Richards and others. This is no place for rewriting their work. I must confine myself to a very general comment. In so far as they were concerned with the history of central banking as such, they were looking at it through the eyes of the inter-war period, and consequently were concerned with the development of the Bank's control of the supply and price of credit — the development, in short, of the bank rate technique. Part of the difficulty of writing this history[7] is that in the nineteenth century central banking development was an empirical, almost unconscious, process, and rationalization of completed processes is easier than exhaustive analysis of policy when it is still in process of formation. In the face of the inevitable obscurity, it is perhaps rash to venture further interpretation than has been given at length in the works of the authors I have named. But, being rash, I venture to suggest that what was happening was (1) that in the middle decades of the nineteenth century one of the most important sensitive quantities in the English economy was mercantile credit; (2) that, as the century wore on, the volume of English long-term foreign lending became an even more important sensitive quantity, influencing the whole world economy; and (3) that these two critical quantities were found by experience to be peculiarly sensitive to the Bank of England bank rate technique. The course of the trade cycle in England depended a great deal in the earlier period, on what was happening to mercantile credit; then later, on the course of English foreign lending at long term. Indeed, in the latter part of the century, it was by controlling, through its bank rate technique, the timing of England's foreign lending, that the Bank of England was able to force the burden of adjustment onto the borrowing countries, and so to maintain an international gold standard with little bother for England itself.

If this is a fair comment on the earlier phase of central banking, it means that the Bank of England discovered that its power lay in touching the economy at particularly sensitive spots. In the light of this, the recent swing away from "traditional central banking" does not appear as such a complete break from the past. With the changing structure of the economic system the sensitive points have

7. A difficulty of which I myself was acutely conscious when, in my *Bank of England Operations, 1890–1914*, I tried to make a contribution.

changed: first there was mercantile credit, then foreign lending, and now perhaps stock exchange speculation or consumer credit. If this is so, it is entirely in keeping with the basic tradition that the central bankers should seek out these newly-important sensitive spots and find weapons by which they can be touched. The central bank should be quick to adapt itself to changes in the economy, and should be ready to use any device it can find to control the behavior of the financial system in the interest of the "employment policy" adopted by the government.

The Macmillan Committee was thus entirely right to open its remarks on the functions of a central bank with a phrase about keeping the financial structure on an even keel. What has perhaps been less far-sighted has been the definition of central banking in terms of the bank rate and open market operations, both of which just happened to be enormously important through a long formative phase. The dangers in thinking of central banking in the narrower terms of bank rate technique, instead of in the broader terms of squeezing the financial structure at sensitive critical points, were seen in the inter-war period, when the crystallization of central banking thought led to numerous attempts to transplant "traditional" central banking into entirely unsuitable environments. The mistake was not in thinking that the establishment of central banks was desirable, but in thinking that the critical points in diverse economies would always be sensitive to a bank rate technique of the traditional English kind.

## VI

The moral of all this can be put very shortly. It is that the cardinal virtue of the central banker is not conservatism in technique, but rather a disposition to discover novelties and to be versatile in technique. We are gradually learning that while the dictator can use the big stick, and laissez faire can rely on the high price, economic planning in a truly democratic society means pressing into service every conceivable device — persuading, cajoling, inciting people, edging the economy now a little in this direction, now a shade in that. If central banks are to play their full part in this process, our central bankers must not be content to stick to the traditional technique. They must sharpen their insight into the working of the economy, so as to see not only where we stand in the trade cycle but also where are the critical sensitive points in the financial system; and they must be quick to adapt their methods to suit the changing economic structure. In two words, they must be percipient and they must be versatile.

R. S. SAYERS.

LONDON SCHOOL OF ECONOMICS

# THE BRITISH PAYMENTS
# AND EXCHANGE CONTROL SYSTEM

## SUMMARY

I. Bretton Woods and convertibility, 212. — II. The Exchange Control Act of 1947, 214. — III. Sterling balances, 216. — IV. Sterling area agreements, 218. — V. The American Account, Transferable Account and Bilateral areas; and their relative importance, 222. — VI. Payments agreements with the Transferable Account area, 226. — VII. And with the Bilateral area, 230. — VIII. "Administrative Transferability," 233. — IX. The intra-European payments agreement and American aid, 234.

The British exchange control system, built up during the war, is now governed by the permanent Exchange Control Act of 1947, by Treasury Orders made under it, and by various payments agreements between Britain and other countries which form negotiated counterparts of the domestic legislation. Similar legislation is in force elsewhere in the sterling area. It is the purpose of this article to give a general analysis and assessment of the Act, orders, and agreements (including the European co-operation agreements) as the system had evolved to its position at the beginning of 1949, and to cover relevant aspects of Britain's exchange position.

## I

The Act, composed mainly of exchange prohibitions and obligations, makes general statutory exceptions in favor of the sterling area (technically called "the scheduled territories") and confers further wide powers of exception, general and specific, on the Treasury. Under it the British exchange resources are collected, conserved, and allocated in the interests of Britain and the area as a whole. The principle of exchange control was accepted more or less by all parties, but the Conservatives voted against its formally unlimited duration.[1] In most matters the Act followed war-time precedents.

In framework it is consistent with British obligations under the Bretton Woods Agreement. For, with its sterling area provisions and Treasury exceptions to the statutory restrictions, it *could* easily be operated to give a Bretton Woods freedom to current transactions while imposing a Bretton Woods obstacle to undesirable capital movements. The protection of sterling by the Act against specula-

1. The Act is 10 and 11 George 6, Ch. 14. For the main debate, see Hansard November 26, 1946, voting being 300 for and 125 against. The Act came into force on October, 1, 1947 when it replaced war-time controls.

tive attack gives sterling greater stability and hence aids world trade, itself a Bretton Woods objective. Moreover, under that agreement, Britain has the right to impose restrictions on current transactions until 1952.[2] As Britain has a natural long-term interest in multilateral trade and in multilateral use of sterling, [3] the controls *would* be most liberally administered with full Bretton Woods freedom for current transactions after the transition. This is done now in respect to the so-called "transferable account" area.[4] But despite that long-term interest in current transactions, British capital movements overseas will need to be controlled in the national interest for many a year, and for this alone the Act is necessary.[5]

Under the American Loan Agreement of 1945,[6] however, Britain agreed with the United States not to impose restrictions on payments and transfers for current transactions from July 1947 "unless in exceptional cases a later date is agreed upon after consultation." Accordingly, arrangements had been made by July with a great many countries to this effect, and negotiations were in train with the remaining countries. The general convertibility collapsed, however, in August, and hurried consultation with the United States took place, ending with an exchange of letters which accepted, as "a temporary and emergency measure," the suspension of general transferability.[7]

2. See Article XIV. All signatories except the United States, Guatemala Mexico, Panama, and Salvador have availed themselves of this Article (*I.M.F Report for 1947*, page 33).

3. "It was not our aim to encourage bilateralism. Quite the contrary: we knew only too well that multilateralism was the only satisfactory basis for our trading. It was for that reason that we were intent on maintaining the largest possible area of multilateral trading on sterling." (Chancellor of the Exchequer, *Board of Trade Journal*, January 1, 1949, page 4. I shall refer to this *Journal* as *BTJ*.) For detailed analyses of British Trade supporting this, view, see articles by G.D.A. Macdougall in the *Economic Journal* for March 1946. and 1947.

4. The sterling and transferable areas cover 46 per cent of the world's population.

5. In the second reading debate the Chancellor declared: "We cannot afford now or in the foreseeable future any large overseas lending." (col. 1426.) On the other hand, aid from the United Kingdom to other countries since the war to the middle of 1948 amounted to £870 million of which £400 million was gift. (Hansard, November 30, 1948, col. 1816.)

6. Cmd. 6968 of December 1945.

7. The I.M.F. doubted the wisdom of the British attempt: "The assumption of the obligation to permit the transfer of current sterling may add significantly to the burden of the deficit in the British balance of payments. There is therefore a danger that, if efforts to improve the United Kingdom's international position and the general international payments situation are not effective, the assumption of this obligation may compel the imposition of even more severe restrictions on imports and thus to some extent reduce the flow of world trade." (*Report*

As a concomitant, the various agreements arranged or in process had to be reviewed and revised. By the summer of 1948 much of this had been done, and although changes are still taking place, a new structure has been built up. Still more recent are the repercussions of the European Economic Cooperation plans and the intra-European payments agreement.

Thus the British exchange system, while remaining constant enough in the domestic sphere, has had a rapidly changing international environment which would appear only now to be settling down. The need for flexibility still remains, and we must expect a varying content in payments agreements for some time to come. But it would appear that more or less stable principles have emerged on current transactions.

## II

The formal structure of the Act and details of its prohibitions may now be taken up. The main purpose is to feed into the Treasury's Exchange Equalization Account all available gold and foreign exchange and to regulate the outflow from that central fund. Since the Treasury's powers of dispensation are very wide, virtually all the following account of the Act must be read with the proviso, "except with Treasury permission," at each point. That is, administrative discretion can be used in given cases.

All dealings in gold coin and gold bullion, and in foreign (nonsterling) currencies or bank credits are prohibited between members of the public, and must be transacted only between the public and authorized dealers or between the latter themselves. These dealers, mostly the banks, handle the bulk of the day-to-day business under delegated powers, act as a channel for the Exchange Equalization Account, and are accountable to the Bank of England and the Treasury for their actions. Mere possession as well as dealing is covered, and accordingly all gold coin and bullion, and all important ("specified") foreign exchange must be surrendered to the authorized dealers on official terms. Travellers' checks and similar documents are also regulated. The requirement to surrender applies to gold or currency in one's ownership abroad. Moreover, debts and rights abroad in foreign currencies or in foreign-held sterling must be fully and promptly collected and enforced, with subsequent surrender of the appropriate proceeds. Equally, exports sanctioned under the Act must not be unduly delayed; and all exports must be paid for normally

*for 1947*, page 37.) For example, the adverse movement in the terms of trade alone cost Britain some £250 million in 1947 and this was worsening in 1948. (Hansard, June 24, 1948, col. 1557.)

within six months and always at prices satisfactory to the national interest, with the foreign receipts surrendered. Delay or under-valuation[8] are offenses. The currency in which exports are payable can be prescribed, and this, with the question of payment for imports, leads on to the payments agreements described below.

Thus, all income received or receivable outside the sterling area is swept up fully and promptly into the central pool of the Exchange Account. The duty of surrender falls not merely on British subjects but on all persons in the United Kingdom or deemed resident in it though actually absent. In this way the official exchange rates are maintained and the central fund is fed. It is agreed on all sides that the mobilization of resources is complete and at official rates. No accusation can be levelled against Britain, as has been levelled against France, that its people have private hoards of gold or exchange, that there are black market deals, or that exports are undervalued.

But unlike war-time measures, the Act does not confer power to requisition foreign securities. The cream had already been taken off in the war,[9] and further powers would be of little importance. Instead, the Act in effect blocks the transfer of securities except within the sterling area, and if transfer is permitted, the seller must surrender the capital proceeds or otherwise satisfy the Treasury as to a mere "swap." Where the securities are on a British register, special duties fall on the registrar to comply with the Act. Where they are not on a British register, the certificates of title must be deposited with authorized depositaries, mostly the banks, who in turn are under special obligations. Other provisions cover substitutions, bearer securities, options, and emigrants' capital. A man can no longer do as he likes with his own but must submit his foreign property and income rights to the national need. The whereabouts and ownership of any capital are known, and its foreign realization becomes as much an in-payment to the Exchange Account as any other receipt.

On outward payments the Act is simple, brief, and sweeping. Except with Treasury permission, no one can make any payment in the United Kingdom to or for the credit of any person resident outside the sterling area directly or indirectly, and no one resident in

8. Hence the British terms of trade are not distorted by undervaluation, as is commonly alleged with the French.

9. British holdings of U.S. and Canadian dollar securities equal only $58 million, including many not readily marketable. (Hansard November 4, 1947, col. 1537.) For the Argentine railways and other South American public utilities, see VII below.

the United Kingdom can make any payment outside the United Kingdom to or for the credit of any person resident outside the sterling area. Nor can anyone pay a person inside the sterling area for receipts or property outside that area. The majestic sweep of these clauses is broken by the mundane and hard-bargaining proviso, "except with Treasury permission." A prime duty of the Treasury, accordingly, is to come to terms with other sovereign states and make payments agreements arranging the actual mode of payment for imports and other British debts. The flow of imports is regulated not by the Exchange Control Act, but by the Export and Import Act. The two Acts are geared to one another and where an import license is granted, the exchange will also be granted. A complicated exchange structure has been built up on payments agreements, dealt with below.

But outflows could occur in other ways. Hence the exportation of any of the following is also prohibited: British or other legal tender money, Treasury bills, postal orders, gold, non-sterling bills of exchange, certificates of title to securities, dividends and interest, life assurance policies, and documents certifying destruction, loss or cancellation of such documents, and valuables carried by a traveler. As the value of some of these arises from their import and use in Britain again, their importation is also forbidden as a further deterrent to their exportation.[1] Personal search and postal inspection are also authorized.

The blocking of any transfer of capital securities has already been mentioned. In addition there can be no new issues for countries outside the sterling area of securities and coupons, bearer or otherwise, nor payment of capital moneys in any form, not even annuities, life policies, or trust settlements, except with permission. Special clauses cover the complicated cases and affairs of subsidiary companies registered abroad but wholly or partly British owned.

## III

It will be seen that in general there is a sharp distinction between the sterling area and the rest of the world. Within the sterling area, private payments can be made freely for any purpose, current or capital, and sterling circulates quite freely or nearly so at will.[2] The sterling area, it may be pointed out, is the largest multilateral

1. Hence "free market" dealings in, e.g. £ notes, are highly speculative and depend virtually on smuggling for their values as the amounts legally importable are very small.
2. "The sterling area may be defined as an area within which exchange control does not operate." (Hansard, November 4, 1947, col. 1538.)

trading system in the world, even if the transferable account area is excluded.

But before details are taken up, something must be said about "accumulated sterling balances"[3] — a euphemism for British short-term debts.

Before the war, London as a great banking center naturally carried vast sums on overseas account. In the main these were, by the middle 1930's, the working balances and monetary and banking reserves of the sterling *bloc* of that time, including its non-British members. A sterling area had grown up naturally during the nineteenth century, unplanned and unrecognized as such, when the gold standard was perhaps more a sterling standard than an international one. The countries concerned naturally found in Britain their main market and source of supply and accordingly banked and kept most of their reserves there. In the breakdown of the 1930's they found it more in their national interests to keep linked with sterling than to link afresh with any other currency, and they formed the changing sterling *bloc* of the period.[4]

By the outbreak of the war the more foreign members had withdrawn themselves and their main balances, and with the war the sterling area in its formal and legal sense emerged. The net external liabilities of London on short-term account were then nearly £500 million, relating mainly to the new sterling area. But during the war Britain ran up extra short-term liabilities to the sterling area and elsewhere which by the end of 1946 amounted to £3,000 million, making a grand total of some £3,500 million. But considering current overseas price-levels, note issues and deposits, one might well guess the appropriate, necessary reserves and working balances in London to be three times the immediate pre-war figure, that is some £1,500 million. If so, some £2,000 million stands out as the accumulated sterling balances to be offset, repaid, funded, or even scaled down as war debt—processes envisaged in the American Loan Agreement of 1945.

But no such general settlements have yet been negotiated,[5] and the balances remain blocked except for specially negotiated releases and offsets under various payments agreements, all *ad hoc*, piecemeal and short-term. Apart from cash holdings, a large amount is held

3. *The Report of the Bank for International Settlements* for 1947 gives an excellent account of the origin and growth of these balances. (pp. 104–112.)

4. See the League of Nations' *International Currency Experience* (1944), Chap. III.

5. Australia and New Zealand have voluntarily surrendered £38 million to date. Their sterling holdings may be about £300 million today.

in the form of Treasury bills at one-half per cent, though some short
— and medium — term Government securities are held by, e.g. Colo-
nial Currency Boards at varying higher rates. Total interest amounts
to about £10 million a year.[6]

These balances are mostly held inside the sterling area and,
being blocked, are exceptions to the otherwise wide transferability
of sterling which is characteristic of the area.

## IV

The scheduled territories which now make up the sterling area
are the United Kingdom, Ireland, Australia, New Zealand, South
Africa, Southern Rhodesia, India, Pakistan, Ceylon, Burma, British
Colonies, Iceland and Irak. At various times in the recent past it
also included Faroe Islands, Egypt, Sudan, Palestine and Trans-
jordan.[7] In various measure all members contribute to and draw
from the central reserves of gold, dollars, and other exchange built
up in the British Exchange Account referred to above, but exceptions
are growing. The Colonies are under more or less direct control from
London as to their disbursements and receipts.[8] The remaining
countries listed are self-governing or otherwise autonomous states
and, apart from mutual interest and understandings with London,
they pursue their own course in the light of their own rights and
interests.[9] Published agreements exist, however, with South Africa,
India, Pakistan, Ceylon, Irak, and Burma.

Under the South African agreement,[1] South Africa took control
of its own exchange receipts and disbursements, withdrew an older
understanding to sell its gold exports for sterling, granted a gold
loan to Britain of £80 million at one half per cent repayable in three
years, and arranged the control and scrutiny of capital movements
from Britain "not required for any useful economic purpose in the

6. Hansard, March 16, 1948, col. 1877, and July 20, 1948, col. 231.
7. Special care has to be taken, therefore, in comparing figures relating to
the sterling area at different dates. Egypt and Sudan were excluded in July
1947; Palestine and Transjordan, in February 1948; the Faroes, in November.
I have kept consistently to the current definition.
8. "In 1947 there was a modest (colonial) surplus with the Western Hemi-
sphere but an overall deficit of not less than £100 million. In 1948 the overall
deficit may disappear or at least be greatly diminished while the surplus with the
Western Hemisphere should be substantially increased." (Hansard, July 29, 1948,
col. 146.) Hong Kong's position, complicated by China, has certain peculiarities
(Hansard, July 20, 1948, col. 24).
9. "The sterling area, like the Commonwealth which is largely co-extensive
with it, is an association of free people and it has to work by mutual trust and
co-operation, not by compulsion," (Hansard, May 14, 1948, col. 2479).
1. Cmd. 7230 of October 1947.

Union." A small purchase program was also included. Thus, South Africa has partly taken itself out of the sterling area. The reference to capital movements suggests a "hot" money flow or capital flight to South Africa,[2] and the whole agreement suggests some monetary or political strain within the area, possibly arising from Britain's greater desire for economy in dollars all around. In any event, South Africa's position has deteriorated, in the main from increased imports, decreasing capital flow, and worsening terms of trade; for gold at a fixed price dominates its exports. South Africa has now drawn, not on the central pool,[3] but on the I.M.F. for dollars, is said to be seeking an American dollar loan, and has had to cut the gold cover for its currency in order to free gold for export.[4]

The Indian and Pakistan agreements[5] illustrate the technique of settling and gradually releasing or offsetting accumulated sterling balances. They were agreed upon at £1,160 million for India as a whole. Under the main agreement, a "No. 1 Account" was to carry sterling currently earned and certain transfers from a "No. 2 Account" which carried the main blocked balances. Sterling in the first account was (originally) made freely available for *current* purposes in any currency area and for *any* purpose in the sterling area. This account was credited with £35 million transferred from the second (blocked) account and with a further £30 million transferred as a working balance. Further debits and credits between the accounts could also arise from military and war matters and from pensions. The net capital movement between India and the rest of the area, as ascertained by the Reserve Bank and the Bank of England, was to be transferred from the blocked account to the current account if favorable to India, and *vice versa* if unfavorable.

Following the partition of India, the independent governments of India and Pakistan came to a mutual agreement as to existing balances, and agreed that, from January 1, 1948, each country would retain its own foreign earnings[6] and be debited with its own foreign

---

2. There had been a capital efflux of £35 million in the year ending June 1947 (Hansard, November 4, 1947, col. 1536). See also the I.M.F's *Financial News Survey* (July 15, 1948 and November 23, 1948), which quotes unofficial estimates of £182 million in 1946–47 and almost £75 million in the first half of 1948. I shall refer to the *Survey* as *FNS*.

3. South Africa holds very little sterling, e.g. £11 million in 1946. (*B.I.S. Report for 1947*, page 106.)

4. *FNS*, October 7, 1948 and November 11, 1948.

5. Cmds. 7195 of August 1947, 7342 and 7343 of February 1948, 7472 and 7479 of July 1948. The earlier agreements were for short periods, but the latest are for three years, subject to consultation if circumstances change radically.

6. India's balance with the United States, which is traditionally favorable, now seems doubtful. (*FNS*, August 26, 1948.)

expenditure; and that each would negotiate separately with the British Government as to further sterling releases. Accordingly, Britain transferred a further £18 million to India's current account,[7] and India agreed not to draw on the central pool for more than £10 million worth of certain hard currencies. Britain also transferred £6 million to Pakistan's current account, together with a further £10 million as a working balance. Pakistan agreed not to draw on its sterling by more than £3.3 million worth of the same hard currencies. The latest agreements free £40 million a year for India from 1949 (hard currencies being £15 million) and up to £12 million for Pakistan from 1948 (hard currencies being £5 million). In addition, £100 million was written off the blocked balances against British military stores transferred, and £176.25 million for commuted pensions now taken over by Britain.[8] These last offsets amount to about a fifth of India's old sterling balances.

The Ceylon agreement[9] recognizes that Ceylon is a net earner of dollars but is on balance a capital debtor to Britain. The sterling assets of Ceylon are to be ascertained, and various No. 1 and No. 2 Accounts are set up. The first carry all currently earned sterling and certain transfers from the second (blocked) accounts amounting to £3.5 million, plus a working balance of some £4 million and certain accrued interests. Net capital movements between Ceylon and the sterling area are to be adjusted between the two kinds of account. Economic development is arranged for. Ceylon's net earnings of American and Canadian dollars, estimated at £5.3 million in 1948, will fall into the central pool.

Three Irak agreements[1] cover complicated arrangements also turning on No. 1 and No. 2. accounts. The first accounts carry all currently earned sterling and a series of transfers from the second (blocked) accounts. These cover £15 million to be transferred over five years, together with the equivalent of certain military stores and railway supplies, £2 million as a working balance, £5 million as a special provision, and various interest and other transfers. Net capital movements are to be debited or credited as the case may be. Originally, the sterling in the first accounts was freely available for

7. This does not seem to have been fully used. (Cmd. 7472 and *FNS*, September 16, 1948.)
8. See also Hansard, July 15, 1948, cols. 1404–1410. Britain has admitted a British war-debt of £55 million as a counter claim.
9. Cmd. 7422 of June 1948.
1. Cmds. 7201 of August 1947, 7269 of November 1947, 7490 of June 1948. Irak at one time placed a partial ban on exports paid in sterling, but latterly its sterling holdings have fallen. (*FNS*, September 9, 1948 and December 23, 1948.)

E

current transactions in any currency area and for any purpose in the sterling area, but subsequent agreements limited non-sterling use to some £5.75 million worth of certain hard currencies in a year and later to £6.25 million worth for the second year.

Apart from a small Burmese agreement[2] granting £2 million in hard currencies in addition to Burma's own earnings, these are the only published agreements within the sterling area. Ireland, however, has certain dollar-drawing rights up to £14 million, plus her own dollar earnings.[3]

All told, the *net* changes in the sterling balances, including for convenience of reference the net changes in the non-sterling area, have been as follows:[4]

NET CHANGES IN STERLING BALANCES (£ MILLION)

|  | 1946 | 1947 | 1948 (first half) |
|---|---|---|---|
| Sterling area...................... | − 37 | − 147 | + 116 |
| Non-sterling area.................. | + 76 | + 2 | − 142 |
|  | + 39 | − 145 | − 26 |

Figures for *gross* changes are not known despite a multiplicity of fragmentary data. It would seem, however, that gross releases and offsets under formal agreements with Argentina, Brazil, Ceylon, Egypt, India, Iraq, Pakistan, and Uruguay alone totalled £156.5 million in 1947 and about £540 million in 1948.[5]

Subject to the important and growing exceptions discussed above and to any informal arrangements inside the sterling area, the British Exchange Account ultimately deals with not merely British receipts and expenditure as regulated by the Exchange Control Act described above but also handles the net foreign receipts and disbursements arising elsewhere in the area. In addition, it received the original American and Canadian loans and the British drawings on the I.M.F., and it currently receives dollars under the Marshall Plan. Its reserves of gold and exchange, however, are not merely the United Kingdom's: they are also the ultimate reserves of the whole sterling area. It is declared policy not to allow them to sink below £500

2. Cmd. 7560 of October 1948.
3. *BTJ,* November 22, 1947, page 2043 and *B.I.S. Report for 1948,* page 96. Ireland probably holds £200 million of sterling.
4. See Cmd. 7520 on U.K. Balance of Payments 1946–48, page 6. The current definition of the sterling area is consistently applied.
5. Hansard, July 5, 1948, col. 20; July 13, 1948, col. 1011; July 15, 1948, col. 1404–10; July 22, 1948, col. 55–57; January 20, 1949, col. 53.

million.  Should they so sink, sterling area exports to the rest of the world would need to be further expanded or imports further curtailed.

The changes in the reserves of gold and American and Canadian dollars are known.  But they cannot be divided up accurately between the United Kingdom itself and the rest of the area.  For example, the dollar expenditure of British oil companies is a United Kingdom charge although much of the output is sent to the rest of the area. Allocation is impossible in dealing with the loss of reserves to countries which convert only part of their sterling receipts into gold and dollars. Subject to these remarks, the *net* changes in the central pool of gold, American and Canadian dollars have been as follows:[6]

NET CHANGES IN GOLD AND DOLLAR RESERVES (£ MILLION)

|  | 1946 | 1947 | 1948 (first half) |
|---|---|---|---|
| On United Kingdom account......... | − 342 | − 657 | − 186 |
| On remaining sterling area.......... | + 45 | − 204 | − 19 |
| On entire sterling area.............. | + 71 | − 163 | − 49 |
|  | −226 | −1024 | − 254 |

Holdings of gold and dollars amounted to £457 million worth on December 31, 1948, to which some £90 million of E.R.P. reimbursements must be added.

## V

For purposes of exchange control, the non-sterling world is divided into the following groups, each with its own characteristics.[7] Payments agreements or other negotiated arrangements exist with the countries marked with an asterisk; for other countries, payments are regulated by Treasury Order only.

A fourth "miscellaneous" group comprises the remainder of the non-sterling world, e.g. Afghanistan, Arabia, and so on.[8]  Trade with Germany has, of course, its own special regime agreed upon by the United States and Britain and operated through their Joint Export-Import agency.[9]

6. Cmd. 7520 of September 1948 (United Kingdom Balance of Payments, 1946–48), page 6. For the second half of 1948, the net change was − £169 million (Hansard January 27, 1949, col. 1123), bringing the total for the year up to −£423 million.

7. Exchange Control (Payments) Orders, 1948, viz., S.I. 1080, 1418, 1736, 1856, 2423 and 2628.

8. I include Paraguay although it is under unique arrangements. (e.g. Hansard November 25, 1947, col. 1791.)

9. Cmds. 7001 of December 1946 and 7301 of December 1947.

| American Account Countries | Transferable Account Countries | Bilateral Countries |
|---|---|---|
| *United States | (a) | (a) |
| Philippines | *Czechoslovakia | *Belgian Monetary Area |
| Bolivia | *Dutch Monetary Area | *Canada and Newfoundland |
| Colombia | *Norway | *Danish Monetary Area |
| Costa Rica | *Sweden | *French Franc Area |
| Cuba | (b) | *Portuguese Monetary Area |
| Dominica | *Chile | *Switzerland |
| Ecuador | *Egypt | (b) |
| Guatemala | Ethiopia | *Argentina |
| Haiti | *Finland | *Austria |
| Honduras | Iran | *Brazil |
| Mexico | *Italy and Trieste | Bulgaria |
| Nicaragua | *Poland | China and Formosa |
| Panama | *Russia | Greece |
| Salvador | Siam | *Hungary |
| Venezuela | *Spanish Monetary Area | Japan |
|  | Sudan | Palestine |
|  |  | *Peru |
|  |  | Roumania |
|  |  | Syria |
|  |  | Tangier |
|  |  | Transjordan |
|  |  | *Turkey |
|  |  | *Uruguay |
|  |  | Vatican City |
|  |  | *Yugoslavia |

British exports to any of the sixteen countries in the American account area are payable in United States dollars or in sterling held by residents in any of them.  Sterling received by any of them on current account is widely transferable and convertible into United States dollars and, therefore, universally transferable.  Under the Loan Agreement of 1945 there is no restriction on the use or convertibility of the United States's current receipts of sterling.  By way of example in other countries, a Bolivian resident can freely transfer his sterling to another Bolivian, or to a resident in any of the fifteen other countries in the area, in any of the fifteen countries in the transferable account area, in the miscellaneous countries, and, of course, in the sterling area itself.  In short, anywhere in the world except in countries in the bilateral area.  If the transfer is to the United States, the sterling becomes convertible into dollars.

Exports to any of the fifteen countries in the transferable account area are payable in locally held sterling, in sterling from other countries in the area, or from the American account area.  Alter-

natively, exports to Czechoslovakia, Dutch territories, Norway, and Sweden may be paid in the respective local currencies. Sterling received by any of the countries is freely transferable within that country, within the whole area on current account only, to the miscellaneous countries, and, of course, to the sterling area. Moreover, specific sanction may be given to still wider transfers. (See below, VIII and IX.)

The "miscellaneous" countries are minor ones in world trade. Exports to them are payable in locally held sterling or in sterling from a transferable or American account. Sterling received by them can be spent among themselves or in the sterling area.

The wide, automatic transferability of sterling implied by the above does not necessarily apply to the bilateral area. Exports to a country in sub-list (a) are payable in its locally held sterling or in its local currency; exports to a country in sub-list (b) are payable in locally held sterling only. Sterling received by each country is transferable only within that country or, of course, to the sterling area. It cannot be transferred outside except with special Treasury permission. (But see below, VIII and IX.)

As a payments agreement with Britain is in effect also an agreement with the rest of the sterling area,[1] the relative importance of the different areas distinguished above could be fully determined only from statistics on the whole area's visible and invisible items. But this information is not available.

Nevertheless a reasonable picture can be formed from Britain's own trade figures, as given in the table on the following page.

A pattern which is slowly changing — slowly, for trade cannot be turned on and off like water in a tap — is discernible here and may be described as follows. Countries in the sterling and transferable areas have been supplying Britain with a rising proportion of its imports, the share of the first being over a third, and of the second, over a fifth. Together they approach 60 per cent. If all E.R.P. countries are added in (see IX below), two thirds of the import sterling is transferable. The sterling paid out is automatically transferable over much of the world and is mopped up mostly by return British exports. The British balance of trade breaks about even, and more than even with the sterling and transferable areas, when freight and insurance are allowed for. A falling fraction of imports

---

1. But there has been only one deal in which Britain, Australia, India, New Zealand, South Africa, and the Colonies have formally acted as one, namely, a sterling deal with Japan (*BTJ*, November 13, 1948, page 942; *FNS*, November 18, 1948). Other countries in the sterling area may accede if they wish.

UNITED KINGDOM IMPORTS (C.I.F.)[2]

| Currency Areas as of January 1, 1949 | January–June 1947 | | July–December 1947 | | January–June 1948 | | July–December 1948 | |
|---|---|---|---|---|---|---|---|---|
| | £m | % | £m | % | £m | % | £m | % |
| Sterling area.......... | 254.2 | 31.0 | 303.8 | 31.2 | 372.2 | 36.3 | 376.7 | 35.8 |
| American area viz:..... | 181.6 | 22.1 | 183.1 | 18.8 | 126.5 | 12.3 | 114.2 | 10.8 |
| *U.S.A...............* | *150.7* | *18.4* | *146.5* | *15.0* | *95.4* | *9.3* | *89.0* | *8.4* |
| *Other...............* | *30.9* | *3.8* | *36.6* | *3.8* | *31.1* | *3.0* | *25.2* | *2.4* |
| Transferable area viz:.. | 106.3 | 13.0 | 155.4 | 16.0 | 198.0 | 19.3 | 230.7 | 21.9 |
| *E.R.P. countries.......* | *55.2* | *6.7* | *88.3* | *9.1* | *93.4* | *9.1* | *117.7* | *11.2* |
| *Other...............* | *51.1* | *6.2* | *67.1* | *6.9* | *104.6* | *10.2* | *113.0* | *10.7* |
| Bilateral area viz:...... | 265.2 | 32.3 | 310.3 | 31.8 | 305.4 | 29.8 | 307.5 | 29.2 |
| *Canada & Newfoundland* | *109.2* | *13.3* | *129.9* | *13.3* | *117.2* | *11.4* | *105.8* | *10.0* |
| *E.R.P. countries.......* | *67.9* | *8.3* | *74.3* | *7.6* | *99.2* | *9.7* | *90.3* | *8.6* |
| *Other...............* | *88.1* | *10.7* | *106.1* | *10.9* | *89.0* | *8.7* | *111.4* | *10.6* |
| Germany............ | 5.6 | 0.7 | 13.7 | 1.4 | 14.5 | 1.4 | 14.9 | 1.4 |
| Miscellaneous area..... | 2.7 | 0.3 | 4.6 | 0.5 | 5.1 | 0.5 | 7.4 | 0.7 |
| Foreign fisheries....... | 4.8 | 0.6 | 3.2 | 0.3 | 4.0 | 0.4 | 2.4 | 0.2 |
| Total.............. | 820.4 | 100.0 | 974.1 | 100.0 | 1025.7 | 100.0 | 1053.8 | 100.0 |

UNITED KINGDOM EXPORTS AND RE-EXPORTS (F.O.B.)[2]

| | £m | % | £m | % | £m | % | £m | % |
|---|---|---|---|---|---|---|---|---|
| Sterling area.......... | 247.6 | 45.0 | 313.7 | 48.5 | 371.9 | 48.0 | 422.6 | 48.4 |
| American area viz:..... | 42.7 | 7.7 | 39.3 | 6.1 | 50.3 | 6.5 | 52.8 | 6.1 |
| *U.S.A...............* | *33.8* | *6.1* | *27.6* | *4.3* | *34.2* | *4.4* | *36.8* | *4.2* |
| *Other...............* | *8.9* | *1.6* | *11.7* | *1.8* | *16.1* | *2.1* | *16.0* | *1.8* |
| Transferable area viz:.. | 105.3 | 19.1 | 107.5 | 16.6 | 137.6 | 17.7 | 161.7 | 18.5 |
| *E.R.P. countries.......* | *53.6* | *9.7* | *58.0* | *9.0* | *73.8* | *9.5* | *92.4* | *10.6* |
| *Other...............* | *51.7* | *9.4* | *49.5* | *7.6* | *63.8* | *8.2* | *69.3* | *7.9* |
| Bilateral area viz:...... | 142.1 | 25.8 | 167.5 | 25.9 | 194.2 | 25.1 | 217.5 | 24.9 |
| *Canada & Newfoundland* | *20.0* | *3.6* | *26.0* | *4.0* | *33.3* | *4.3* | *40.4* | *4.6* |
| *E.R.P. countries.......* | *76.8* | *14.0* | *87.1* | *13.5* | *105.1* | *13.6* | *115.8* | *13.3* |
| *Other...............* | *45.3* | *8.2* | *54.4* | *8.4* | *55.8* | *7.2* | *61.3* | *7.0* |
| Germany............ | 10.8 | 2.0 | 17.2 | 2.7 | 17.2 | 2.2 | 13.9 | 1.6 |
| Miscellaneous area..... | 2.2 | 0.4 | 2.2 | 0.3 | 3.6 | 0.5 | 4.4 | 0.5 |
| Total.............. | 550.7 | 100.0 | 647.4 | 100.0 | 774.8 | 100.0 | 872.9 | 100.0 |

2. Compiled from the *Monthly Trade Accounts* which are subject to correction in detail. Imports exclude fissionable material and include German reparation goods. Exports include relief and rehabilitation supplies (U.N.R.R.A. and others). Because imports are valued c.i.f. and exports f.o.b., when goods are carried in British ships or under British insurance, the exports would be greater and the imports less in currency than the official values, probably by more than 10 per cent on the average. That is, the trade gap is less than the official figures show as above, subject to an allowance for exports mopping up old sterling balances.

is obtained from Canada, Newfoundland, and the American account area, the first showing a slow drop to about a tenth, and the second a rapid drop to almost the same figure. But here British trade is hopelessly out of balance, save for the generous Marshall and Canadian aid. Countries in the bilateral area (excluding Canada) supply rather less than 20 per cent of the imports. Since their sterling is not automatically transferable, their balances are not additive, and trade with the sterling area outside Britain is a special complicating factor in some cases. Britain alone has a favorable balance with eleven of them, and an unfavorable with ten. The final balancing of the separate accounts requires invisibles, the trade of the rest of the sterling area, the use of credit limits, sale of capital assets, or payments in gold or dollars. (But see below, VIII and IX.)

All told, British exports and re-exports have recently been equalling British imports in f.o.b. values. This is an amazing achievement: before the war, in 1938, they were less than two-thirds. Such a balancing is now Britain's fate, for the old pre-war net invisible income seems gone forever.[3]

## VI

The general characteristics of the different areas have already been indicated, but more insight into the nature of the system is given by a study of the various agreements negotiated with foreign countries.

The provisions of the Anglo-American Loan agreement are well known.[4] The other countries in the American account area are so associated with the United States that, as with the latter, all business must be conducted on a dollar or near dollar basis, and no agreements exist with them. The main imports involved are sugar from the Caribbean, tin from Bolivia, and oil from Mexico and Venezuela.

About thirty agreements are in force with countries in the transferable area. The core is the transferability clause *reciprocally agreed*. For example, from the Chilean agreement:[5]

3. "For the future, in fact it will only be by the export of goods that we shall be able to pay for the goods we import," *Economic Survey for 1948* (Cmd. 7344), page 13. Despite a population growth of some three million, British exports at the end of 1948 were about 50 per cent above pre-war volume, and imports, 20 per cent below — a double sacrifice in goods foregone and goods not obtained. In this balance the British achievement is unique.

4. Cmd. 6968 of December 1945.

5. Cmd. 7497 of June 1948. The Chilean government agrees not to restrict the acceptance of sterling.

"The Government of the United Kingdom shall not restrict the availability of sterling standing to the credit of any Chilean Transferable Account for making —

(a) transfers to other residents of Chile;

(b) payments to residents of the Scheduled Territories; or

(c) payments in respect of current transactions to residents of countries outside Chile and the Scheduled Territories as may be agreed between the contracting Governments."

In the Chilean case the countries in (c) happen to be listed in an appendix as the transferable countries, but the agreement is not necessarily limited to them. All that the listing means is that transfers in such cases are automatic and the formal agreement still remains such transactions "as may be agreed." In given cases, wider transferability may be sanctioned by special administrative agreement. This is standard flexibility. (See below, VIII.) The rates of exchange are based on dollar cross-rates, and all payments between Chile and the sterling area must be settled in sterling, with facilities for remittances "in no way less favorable" than those accorded to other countries. Trade in copper, nitrate, and iodine is subject to "special arrangements" not specified.[6]

The Czechoslovakian agreements[7] name the rate of exchange, allow trade in crowns and sterling, control capital movements "which do not serve direct and useful economic or commercial purposes," and fix mutual credits (£1 million on the British side and rather more on the Czech to cover some old Czech debts) beyond which, however, gold is payable. Seven agreements with the Netherlands[8] settle the rates of exchange and currencies with the Dutch monetary area, control capital movements, and fix mutual credits (£10 million on the British side and, through a war-time debt settlement, possibly up to £55 million on the Dutch), beyond which gold is payable. Complicated and, one fears, acrimonious negotiations are embodied in two agreements which followed Egypt's departure from the sterling area.[9] Egypt's sterling balances were to be ascertained. No. 1

6. *FNS* (July 13, 1948) reports copper as payable in dollars.

7. Cmd. 6694 of November 1945, 7174 of July 1947, and 7585 of November 1948.

8. Cmds. 6681 of September 1945, 6921 of September 1946, 7051 of February 1947, 7358 and 7381 of March 1948, 7487 of June 1948, and 7531 of September 1948. These agreements produced a piece of parliamentary doggerel parodying Canning's Dutch negotiations a century ago:

> "In matters of claims the fault of Cripps
> Is giving too freely and asking for nix."

(Hansard, April 13, 1948, col. 787.)

9. Cmds. 7163 of June 1947 and 7305 of January 1948. They contain references to "arduous labours" in negotiations and to "genuine" trade, with

and No. 2 accounts were set up, the first to carry currently earned sterling, and the second, the blocked balances. A complicated set of releases and transfers was arranged between the two accounts together with certain offsets arising from old British military supplies and stores, provision of hard currencies, and certain capital movements. Unlike other transferable clauses, the Egyptian clause does not sanction automatic capital transfers to the sterling area but only those specially agreed upon.[1]

The Finnish agreement[2] places all transactions on a sterling basis through dollar cross-rates and, although not containing the standard transferability clause, arranges for "necessary flexibility according to circumstances," a phrase operating to the same effect. The Italian agreement,[3] in placing transactions on a sterling basis, fixes the rates of exchange through "corresponding effective rates" for the dollar and limits the wider transferable use of sterling to sterling "at the disposal of the Ufficio Italiano dei Cambi" — clauses no doubt reflecting the complexities of the Italian situation. Three Norwegian agreements[4] settle the rate of exchange, allow trade in krone and sterling, control capital movements, contain the essence of the transferability clause, and provide for mutual credits (£5 million).

The Polish agreement[5] is in standard current form, transactions being in sterling and rates on dollar cross-rates. The two Russian agreements[6] are more complicated for they contain trade transactions and some war-time debt settlements as well as exchange arrangements. The actual trade transactions relate to a deal in

a zig-zag of qualifying letters. Egypt's exclusion from the sterling area was due, one may presume, to its inability or unwillingness to exercise adequate control over its residents in exchange matters. (See *BTJ*, July 12, 1947, page 1198.) Its sterling balances were £404 million at the end of 1945 and £368 million in March 1948. (Hansard, July 20, 1948, col. 19.)

1. A capital movement between Egypt and Irak in sterling has been mooted. Egypt has tended to try for barter agreements with various countries while at the same time not fully using its disposable sterling. (*FNS*, October 7, 1948 and October 14, 1948.)

2. Cmd. 7166 of July 1947. The agreement formally allows full availability for sterling "at the disposal of the Bank of Finland" for current purposes, but in practice it would seem to be transferable sterling only.

3. Cmd. 7587 of November 1948. Italy seems to have a small deficit with the sterling area (*FNS*, November 18, 1948). Vatican City is not included.

4. Cmds. 6697 of November 1945, 7162 of June 1947, and 7474 of July 1948. Norway seems to have a small surplus on current account. (*FNS*, August 19, 1948.) The agreement to pay gold above the credit limit, and the Norwegian obligation to keep a minimum balance of £20 million have recently been waived. (*BTJ*, December 25, 1948, page 1205.)

5. Cmd. 7352 of March 1948. An important trade deal has been added.

6. Cmds. 7297 and 7439 of December 1947. For an important debate, see Hansard, January 27, 1948.

coarse grains against light rails; other transactions are envisaged and put down for future negotiations which, however, seem to be hanging fire. As part of the grain deal, some war debts have been re-arranged on longer terms and much lower interest rates. No prices or exchange rates are mentioned, and the deals are essentially barter, but accounts are to be settled in sterling with standard transferability.[7] The main Spanish agreement[8] follows standard form, including the sterling basis for all transactions and orderly cross-rates, but in regard to the use of sterling in the transferable area, it refers to "direct" current transactions only. Three Swedish agreements[9] name and change the exchange rate, include the use of krone and sterling, control capital movements, and arrange credit limits. (£8 million, but see below.)

London's pre-war position as a holder of working balances and banking reserves appears, though faintly, in these agreements. This is implied in the seven cases where all transactions must be settled in sterling, to which may be added Ethiopia, Iran, Siam, and the Sudan.[1] The Czech and Dutch agreements provide for mutual minimum balances but the provisions for realization under I.M.F. conditions suggest an excess of sterling holdings. The Norwegian agreement arranges for mutual balances which, when revealed recently, were greater in krone than in sterling, but were also realizable originally under I.M.F. conditions. The Swedish, in addition to mutual balances, provided for a minimum Swedish No. 2 holding of £25 million, but as Sweden is not a member of the I.M.F., all were covered by reciprocal exchange guarantees.

It will be seen that membership in the transferable area does not permit a member to make capital payments in sterling to another member and, in practice, it does not permit them to extract gold or dollars from Britain. They must accept sterling offered by another member or by the sterling area, and they are free, so far as the British control is concerned, to accept American account sterling.

7. Russia does not seem to have taken up fully the agreed British counterpart in goods but is using the balance of sterling to buy wool and rubber in the sterling area. A similar thing seems to be happening with its dollar earnings. (*FNS*, September 16, 1948 and November 11, 1948.)

8. Cmd. 7596 of December 1948. Spain has a small favorable balance with the sterling area. (Hansard, March 26, 1948, col. 2105.)

9. Cmds. 6604 of March 1945, 7170 of July 1947, and 7259 of November 1947. Gold beyond the credit limit is no longer payable; some blocked balances are to be released to finance British trade and third countries. (*BTJ*, December 25, 1948, page 1222.)

1. The official Iranian holdings are guaranteed in gold value. (Hansard, November 24, 1947, col. 1788.)

In special cases and with intent to give sterling a still greater multi-lateral use, transfers with the dollar and bilateral areas may be sanctioned administratively. (See VIII and also IX below.) But on the whole the transferable area is a vast self-balancing system with the sterling area.

## VII

It is generally different with the bilateral area, certainly with its chief members; and the payments agreements are shaped accordingly. But the intent is to expand trade to the maximum possible, subject to no (great) loss of gold or dollars from the central reserves of the sterling area; and almost continuous negotiations are arranging for this expansion on the basis of estimated balances of payments. Bilateral trade is better than no trade. (See also VIII and IX below.)

Three agreements with Argentina[2] embody trade and railway transactions as well as payment arrangements — hard bargains, with Britain in the weaker position. Originally, the sterling received by Argentina was fully convertible and any balances held were covered by a gold guarantee, a guarantee still existing. The old sterling balances, mostly accumulated during the war, were wiped out by the sale of British railways (£150 million). Subsequently full convertibility was withdrawn, and transactions put on a bilateral sterling basis; the meat contract (£100 million) was paid in advance by Britain, together with various additions to original contract prices; and undertakings were given to facilitate supplies of coal, steel, machinery, chemicals, petroleum, and so on. The Austrian agreement[3] is colored and limited by the fact that a full peace treaty with Austria is not yet signed. The Belgian agreement[4] fixes the exchange rate, allows trade in francs and sterling between the two monetary areas, controls capital movements, and arranges mutual credit limits (£12 million on the British side, £12 million on the Belgian, plus £15 million arising from certain outstanding Belgian liabilities), with holdings being subject to reciprocal exchange guarantees and I.M.F. conditions. Beyond the credit limits, gold is payable although "as the National Bank of Belgium is desirous of utilizing sterling

2. Cmds. 6953 of September 1946, 7405 of February 1947, and 7346 of February 1948. Contracts are fixed annually. The first agreement, signed in Buenos Aires, is singularly abrupt in its language but the third is customarily and more courteously worded. There have been various prohibitions on British goods and one refusal to remit financial services; Argentina is currently defaulting on the meat contract and imperiling British rations.
3. Cmd. 6891 of August 1946.
4. Cmd. 7264 of November 1947, abrogating a full convertibility agreement of a few months earlier.

balances as part of its reserves," Belgium is not to claim gold unless it needs it for current transactions.  Belgium has been accepting sterling from countries outside the sterling area.  Britain having lost gold in this and other ways, a Joint Committee has been set up to watch the position and raise trade to a high but balanced level.[5] It is also to be hoped that the intra-European agreement will help here.

The Brazilian agreement[6] places transactions on a sterling basis covers prospective trade at the highest possible level, unblocks some sterling balances and has a transferability clause in standard wording. The clause ("as may be agreed") has not yet acquired an automatic character.  (See, however, the Argentine transfer mentioned below.) The accumulated balances, which were £61.5 million in 1947 and £50 million by mid-1948, are covered in general by a revaluation clause; £10 million are to be released over four years, and more (unstated) as and when Brazil repatriates sterling debts, takes over public utilities, and builds a London embassy.[7]  The Canadian agreement[8] is primarily the loan agreement, parallel to the original American line of credit.  The Danish agreements[9] are unique in having no formal limits to mutual credits, although any holdings are, by the text, convertible into gold.  Under a recent trade agreement, involving butter and bacon against coal, steel, oil, and manufactured goods at new high levels, the Danes have stated their intention to pay off sterling debts.[1]

Six French agreements[2] are very complicated documents. Tumbling out pell-mell fashion in confused and changing French conditions, they cover not only a payments agreement on a franc and sterling basis between the two monetary areas but also wartime debt settlements, a series of large British loans and credits (subsequently funded in part at £100 million), and certain gold repayments and requisitioning of securities in partial discharge of the British loans.  Provision is made for the mutual disclosure of overseas

5. See Hansard, March 16, 1948, col. 1887–91, for supplementary details including provisions for possible re-purchase of the gold and for the drastic checking of Belgium's sterling inflow from third countries.  The Chancellor said: "The only alternative [to this agreement] was a complete cessation of trade between the two countries."

6. Cmd. 7438 of May 1948.

7. Brazil's sterling debt is thought to be about £80 million.  (*FNS*, October 7, 1948.)

8. Cmd. 6904 of March 1946.

9. Cmds. 6671 of August 1945 and 7592 of November 1948.

1. *BTJ*, September 25, 1948, page 603.

2. Cmds. 6613 of March 1945, 6809 of April 1946, 6988 of December 1946, 7112 of April 1947, 7408 of April 1948 and 7430 of June 1948.

assets[3] held by nationals, and capital movements are brought under surveillance. The detailed working of the agreements appears to have presented some difficulties. A qualification no doubt reflecting some of the complexities of the French situation is the British Treasury's Order to confine transfers to No. 1 Accounts of the French government or of *banques agréées*. A Hungarian agreement[4] formally re-opens trade between Hungary and the sterling area, arranges cross-rates with a sterling account kept in London, and re-affirms Hungary's obligations in external debts and similar matters. The Peruvian agreement[5] closely follows the Chilean, including the remittance clause, but the transferable clause though formally equally wide is in practice mostly bilateral, and any sterling holdings are covered by a revaluation guarantee.

Four Portuguese agreements[6] fix the rate and the currencies between the two monetary areas, settle minimum balances with rights to repayment in gold, keep capital movements under control, and arrange mutual credits (£5 million) beyond which gold is payable, as is happening on the British side. The Swiss agreement[7] textually akin to the Portuguese, provides for mutual credits of £5 million plus a Swiss credit "as may be agreed in the light of the estimated balance of payments between the sterling area and Switzerland," a helpful clause which has not, however, prevented Britain from losing much gold at times. The Turkish agreement[8], which includes some trade provisions, operates through Turkish accounts recognized as such by the Bank of England, and, as far as local conditions permit, all payments are to take place at published rates subject to Turkish premiums applied to certain operations.

The Uruguayan agreement[9] is a sterling agreement with orderly

3. Compare the U.S. action.
4. Cmd. 6915 of September 1946.
5. Cmd. 7498 of July 1948.
6. Cmds. 6798 of April 1946, 7050 of February 1947, 7302 of January 1948, and 7401 of April 1948.
7. Cmd. 6756 of March 1946.
8. Cmd. 6907 of May 1945. It has some odd phrases, reflecting perhaps Oriental difficulties, e.g. the Turkish government undertakes to do its best "to secure that payments for goods imported from the United Kingdom to Turkey can be made in accordance with the terms agreed upon between the buyers and sellers in so far as the import of the goods has been authorized." Turkish sterling seems to be quite small and is falling; Turkey has been trying to sell widely in sterling but has had at times to reduce import licenses for British goods. (*FNS*, September 23, 1948, and November 25, 1948.) See also IX below.
9. Cmd. 7172 of July 1947. Uruguay is reported to have prohibited exports of wool to the sterling area and to have sought dollar markets for them (*FNS*, October 21, 1948; it seems to be short of sterling. (*BTJ*, October 2, 1948, page 540.)

cross-rates. Although formally very wide, with transferability "to any country," the sterling is in practice mostly bilateral; accumulated balances (£17 million in 1947) are covered by a gold clause and are to be released as and when public utilities are bought. Some are to be released for building a London embassy and — a unique and friendly point — "for permanent bursaries for use in the United Kingdom." Finally, to close this complicated catalogue so far removed from the sweet simplicities of multilateralism, are the Yugoslavian agreements.[1] They embody barter deals and compensation terms for British property affected by nationalization (£4.5 million) or by enemy occupation.

## VIII

It would be wrong to leave the impression that the bilateralism of Britain was crude and rigid. Everywhere it can be relaxed by administrative transferability. Exactly how great administrative transfers are currently is not known although in the recent past they have been considerable. But the Bank for International Settlements declares that "in practice, a large volume of business has been and continues to be transacted on these lines," and it warns us that their importance ought not to be underestimated in number, scope or aggregate value.[2]

Examples can be picked up from various sources. By the Argentine agreement, £10 million was transferred to Brazil and, more recently, Argentina has paid Hungary in sterling. Syria has required Egypt and Irak to pay in sterling. Poland and Turkey settle part of their accounts in sterling as well as in dollars. In trade agreements between Italy and Denmark, sterling has been used at least as the unit of account. The Chilean agreement makes specific reference to administrative transfers in addition to its standard transferability clause. Recent British purchases from Hungary have been paid in sterling which Hungary can use as it pleases. Uruguay sells sterling to Italy. Official press releases announce "facilities for the use of sterling in third countries" for Brazil, Denmark, Peru and Sweden.[3] This is bilateralism tempered by administrative discretion.

Moreover, out of the twenty-four countries formally in the bilateral area, eleven ranging alphabetically from Bulgaria to Yugo-

1. Cmds. 7600–02 of December 1948.
2. *B.I.S. Report for 1948*, page 150.
3. See Cmd. 6953; *FNS*, September 23, 1948, October 28, 1948, August 19, 1948, July 18, 1948; Cmd. 7497; *BTJ*, January 1, 1949, page 20, October 2, 1948, page 640, May 29, 1948, page 1059, September 25, 1948, page 603, and July 31, 1948, page 228.

slavia have obvious peculiarities in local political or economic conditions which require special handling. Special relations too, but of a very friendly sort, exist with Canada. The great bilateral problems have a much smaller area than the formal area would suggest. Argentina now presents the biggest. In addition to "administrative transferability", has now come with the E.R.P. countries, multilateral compensation.

## IX

Lying unconformably over all areas is the recent agreement for intra-European payments and compensations, devised as part of European economic co-operation under Marshall aid.[4]  The intra-European agreement does not abrogate existing agreements, British or otherwise, but so long as "conditional (dollar) aid" is granted, it modifies their working.  In 1947 the E.R.P. countries and their associated monetary areas had 46 payments agreements *inter se* (e.g. Austria had five such links, Belgium eleven, Britain nine, and so on), with aggregate debit balances of $490 million. Being bilateral, they helped up to a point and then hindered or distorted trade, the disadvantages tending towards the end of 1947 to outstrip the advantages. A small "circuit" clearing was set up between Belgium, France, Italy, and the Netherlands, with other E.R.P. countries (including Britain) as occasional members, but little came of it all, and little could come on so small a circuit.  When the O.E.E.C. was set up in April 1948, all members pledged themselves "to continue the efforts already initiated to achieve as soon as possible a multilateral system of payments among themselves," and the present intra-European agreement is the outcome of this.

But no country can afford to hold an indefinitely large amount of another's currency or make indefinitely large gifts, nor ought any country to seek to impose in this way on others.  For a time it looked as if Britain and the sterling area, first in their own name and second through Belgian accumulations, were to be cast as a Fairy Godmother. Since the rôle would have cost Britain both gold and more austerity, it was not acceptable.  The final scheme embodies, therefore, in

4. The main documents are the General and Technical Reports of the Committee of European Economic Co-operation and the Reports on Payments 'Agreements (1947), the United Nations' *Survey of the Economic Situation and Prospects of Europe* (1948), *Convention for European Economic Co-operation* (Cmd. 7388), *Agreement for Intra-European Payments and Compensations* (Cmd. 7546) 1948, and O.E.E.C., *Interim Report on European Recovery Plan* (December 1948).  See also Polk and Patterson, "The Emerging Pattern of Bilateralism," this *Journal*, November 1947 and R.F. Mikesell, "Regional Multilateral Payments Arrangements," *ibid.*, August 1948.

addition to conditional (dollar) aid, certain principles of commercial policy applicable to "net creditor countries," "net debtor countries," and "intermediate countries" respectively. All should seek to increase their trade, and especially they should maintain the normal exports necessary to the recovery of other participating countries. But net debtors and intermediate countries should do their best to increase their current exports and so reduce the prevailing disequilibrium, and make it less persistent. Net debtors should also in their external expenditure exercise the maximum economy compatible with recovery. Net creditors and intermediate countries should buy from net debtors "additional goods and services as freely as is reasonably practicable," also to reduce the disequilibrium. The accounts, in short, can really be settled only in goods and services, and policies to do this must be pursued as part of the running of the scheme. Special attempts should be made to supply and obtain in Europe goods otherwise only obtainable for dollars.

It was subject to the effective practice of these policies by E.R.P. countries that Britain accepted the scheme. It would insure the minimum loss of the British gold or dollar earnings and reserves which are so necessary to bring that other great gap into balance — its gap, perhaps persistent, with the Western Hemisphere.

Under the scheme, the countries estimated to be creditors on current account in the year ending June 1949 give "drawing rights" in their own currencies to their respective debtors, "equivalent to the United States dollar value of goods and services to be provided by E.C.A. for the purposes" of the scheme. If no such conditional aid is given, no drawing rights are given — the case with Portugal and Switzerland. The rights are gifts because the aid is a gift, the one in currency and the other in dollars, a transmutation effected by Marshallian alchemy.

In this sense Britain has agreed to give sterling drawing rights equivalent to $312 million over the current fiscal year (France receiving $200 million worth, Austria, Germany, Greece and Italy receiving the balance) and receive $30 million from Belgium. The agreement, however, provides that drawing rights (and so return dollars) should be struck "after taking into account the agreed existing resources" of the debtors. Hence, where, e.g. sterling balances exist, these must run down before dollar compensation can be claimed. Accordingly, Britain will let these be drawn on up to $209 million worth, covering Greece, Italy, the Netherlands, Sweden and Turkey. In these two ways the total estimated net deficit of the other participating countries with the sterling area, equal to

$491 million, will be financed this year. As a separate deal, Britain is giving Turkey $8 million worth of drawing rights.[5]

But the above by itself is hardly more than an increase of credit limits found in all bilateral agreements. The second great reform, accordingly, is to extend the circuit clearings to all E.R.P. members.[6] The members accept automatically all "first category compensations" as worked out monthly by the Bank for International Settlements. That is, they accept any operation which will decrease one or more debit balances against an equivalent decrease in one or more credit balances within the whole system and not merely within one small circuit, as previously within Belgium, France, Italy, and the Netherlands. Hence a multilateral flexibility is given to all drawing rights and credit limits. Members, too, must accept any operation which offsets, by means of the drawing rights, any monthly deficit between parties plus balances carried forward from month to month. If the O.E.E.C. estimates of the balances on current account prove substantially correct (and that depends partly on the accuracy of the nationally estimated figures), and if commercial policy is practiced as agreed, the E.R.P. countries and their overseas monetary areas should in effect be a multilateral area, where debts incurred in one currency can be offset against receipts due in another currency, and where gold flows should be small. Where gold is more directly involved, as with Switzerland and Portugal, the members do not accept automatic compensations but require their specific consent, a consent not lightly given.

In this way sterling and other E.R.P. currencies have become widely transferable. The conventional classification of countries into a "bilateral area" so far followed is becoming out of date. But the intra-European scheme is still very new, its commercial policies are still to be tried out, its estimates are still to be verified. The policies will require mobility of labor and changes in industrial structure which may be difficult to achieve and may meet with social resistance. National estimates which result in drawing rights that are gifts to others may be colored by this bare fact. The whole scheme turns on American "conditional aid." Gold is still flowing, e.g. from Britain to Belgium. Europe is still not out of the woods.

But sterling which a year ago was almost a drug on the market,

---

5. Hansard, September 21, 1948, col. 53 and January 27, 1949, col. 1115.
6. Iceland and Ireland are part of the sterling area, Luxembourg is linked with Belgium; Switzerland and Portugal receive no conditional aid. Hence the membership for this purpose falls formally to twelve areas.

discredited and sold short on the collapse of convertibility, is back again in the first rank of world currencies, carefully nursed by the exchange control system, by widening payments agreements, by expanding exports, and an austerity of imports. Over much of the world it is a hard and desirable currency and increasing in strength.[7]

<div align="right">H. A. SHANNON.</div>

UNIVERSITY OF BRISTOL, ENGLAND

7. The phrase is used by the Chancellor of the Exchequer, Hansard, January 25, 1949, col. 93. See also "Harder Sterling," *Economist*, December 18, 1948, pages 1026–27.

# THE A & P CASE:
## A STUDY IN APPLIED ECONOMIC THEORY[1]

"Lawsuits are contests in which prosecution and defense both play to win, and in this process there is danger that no one will become fully aware of the true issues of public policy." Corwin D. Edwards, "An Appraisal of the Antitrust Laws," *American Economic Review Supplement*, May 1946, p. 187.

### SUMMARY

Introduction, 238. — I. Selling Policy, 239. — II. Integration, 244. — III. Buying Policy, 247. — IV. Conclusions, 255.

The Antitrust Division of the United States Department of Justice stands today at the height of its power and prestige. Whether we regard the money available to it, the number of cases it is pre-paring, or — most important of all — the sweeping latitude which recent judicial decisions have given to its actions, it would appear that the gains of the late 1930's were not lost during the war, and that further important advances have been made in the post-war period.

The wider the scope of its work, the more exacting become the demands upon its staff.[2] In particular, the quality of its economic thinking becomes of ever greater importance. This thinking is found primarily in the actual cases which the Division prosecutes. Thurman Arnold, who has by far the best claim to the title of being the father of modern antitrust policy, once remarked: "A dog talks by barking,

1. This is an expanded version of a paper delivered before the 61st Annual Meeting of the American Economic Association, and it appears by courtesy of the past President of the Association, Joseph A. Schumpeter, and the Secretary, James Washington Bell. It is based on the last chapter of an unpublished doctoral thesis, *The Dominant Firm, with Special Reference to the A & P Tea Co.* (Harvard, 1948); cited hereafter as *DF*. The award of a Social Science Research Council fellowship made the investigation possible. Access to the A & P record was afforded by Holmes Baldridge of the Antitrust Division, U. S. Department of Justice; to the exhibits, by James V. Hayes of Donovan, Leisure, Newton, Lumbard & Irvine. Their courtesy and the complete freedom they afforded have earned my sincere gratitude. In discussing the policy involved in this case, I have become deeply indebted to Robert R. Bowie, Milton Katz, and Edward S. Mason. The paper has also benefited from the suggestions of Robert L. Bishop. But none of them is necessarily in agreement with any statement made here, nor responsible for any errors of fact or interpretation.

2. Some details of the economic staff of the Antitrust Division of the Department of Justice for the fiscal year 1948 may be found in *The Budget of the United States Government for the Fiscal Year ending June 30, 1950* (Washington: Government Printing Office, 1949), p. 658.

and we [the Division] talk through litigation." In exploring and appraising the collective mind of the Division, then, it seems best to go case by case.

The recent successful action against the New York Great Atlantic & Pacific Tea Company (mercifully abbreviated to "A & P") affords a particularly good opportunity for such exploration and appraisal. The case was brought in 1944, tried the following year, decided in the District Court in 1946,[3] and immediately appealed. The Seventh Circuit Court of Appeals upheld the conviction at the end of February 1949,[4] and a month later the Company announced that it would not appeal further. Thus the ideas of the Antitrust Division as developed in this case have prevailed, and will influence public policy and business conduct. They deserve examination. The frequent references to this case in the important new book, *Maintaining Competition*, by Corwin D. Edwards, leave us no doubt on this score.[5]

The scope of this article is limited. It is not my purpose to comment on the decisions from a legal point of view; nor to discuss the facts of the defendants' operations, except incidentally. A further aspect of major importance, namely the Department's use of statistical materials, will not be touched upon at all. This paper will accept the facts as stated by the Department of Justice, and have for its purpose an examination into the economic theory or theories by which these facts are given meaning, and the public policy which underlies the action. The economic theory of the prosecution and its implied economic policy are best explored in relation to the business methods of the A & P Company in (1) selling goods, (2) integration of manufacturing and retailing, and (3) buying.

## I. SELLING POLICY

A few matters of terminology must first be explained. Following the usual practice in retailing, "gross profit" is taken as the difference between the cost of goods purchased by the retailer and the total gross revenues of these goods when they are sold. "Expenses" include all costs except the cost of the goods. Since totals are almost meaningless except in relation to sales, it is usual to divide both gross profit, or total margin, and total expenses by total sales; and to think and speak in terms of *gross profit rate*, or *gross margin*, and

---

3. U. S. v. The New York Great A & P Tea Co., 67 Federal Supplement 626 (E. D. Ill. 1946).

4. Commerce Clearing House, *Trade Regulation Reporter: Court Decisions*, Paragraph 62,375.

5. Corwin D. Edwards, *Maintaining Competition* (New York: McGraw-Hill Book Co., 1949). See the index, pp. 321 and 330, for the many references.

*expense rate*, both expressed as percentages of sales.  It will be helpful, and analytically correct, to think of gross margin as the *price* of retailing service, and of expense rate as the *cost* of rendering that service.  The difference, or *net profit rate*, is obviously the net return to the retailer.

Figure I indicates certain matters on which there is agreement between both parties: that the lower the gross margin, the larger the sales; and the larger the sales, the lower the expense rate.  At sales of *OA*, expenses per unit sold would be *AE*, gross margin would be *AM*, profit per unit would be *EM*, and total profit would be the
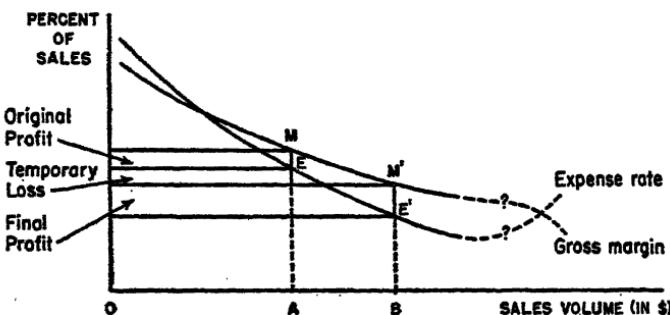


FIGURE I

rectangle labeled "original profit."  Now if the company, in pursuit of greater volume, lowered the gross margin to *BM'*, and *if* consumers for a time did not respond, sales would be no larger, the expense rate no lower, and the profit might be turned into a loss, as represented by the rectangle labeled "temporary loss."  If and when consumers responded, the sales might increase to *OB*, the expense rate fall to *BE'*, and operations again be profitable, as indicated by the rectangle labeled "final profit."  (No attention should be paid to the size of the profit and loss rectangles; only their existence is of interest.)  For example, in experimenting with reduced prices in the last three quarters of 1937, the Company was able to lower its gross margin from 18.6 to 18.3 per cent, to increase weekly sales per store from $1104 to $1226, to lower the expense rate from 18.4 to 17.2 per cent, and hence to increase the net profit rate from 0.18 to 1.22 per cent.[6]

According to traditional economic analysis, unrestrained competition for additional business will drive firms in just this manner to reduce prices and costs, thereby passing on to society the benefits

6. *DF*, pp. 155–158.

of more efficient operation. And the usual attack on monopoly, or monopolistic competition, is that the firm or firms seek greater profit by restricting output short of the least cost point and hence keeping prices up.

But the case against A & P is precisely the contrary: it attacks the Company's attempt to move in the direction of greater output at lower cost. By lowering the gross profit rate in order to induce greater volume, it contends, A & P was selling below cost and unfairly burdening competition. In the words of the Government brief:

> In speaking of how meat business . . . had increased from $200 to $1200 per store, he [John Hartford, the key defendant] pointed out: "This was accomplished by reducing the gross profit rate until the volume was built up to a point where the expense rate was low enough to permit the store to operate at a profit." *We know of no more clear and concise words with which to express the Government's charge* . . .
>
> Hartford then attempted to advance what amounted to a plea in confession and avoidance: "Well, I knew that just as soon as we could get on a proper operating basis, we would have every reason to believe, from our experience, we would get the volume that would take care of it." . . . The evil, however, which is inherent in this pattern lies in the selection of an arbitrary gross profit rate chosen without regard to the expense rate and fixed at a figure which defendants believe will produce the chosen figure of desired volume.[7]
>
> . . . Gross profit rates in selected areas have been set below the expense rates in such areas. [It is incorrect to] argue merely that *eventually* such a practice will result in a profitable operation since the increased volume produced by setting the gross profit rate below the expense rate will *eventually* reduce the expense rates to a point where *ultimately* a profit will be earned.[8]
>
> Of course, the lowering of gross profit rates may *ultimately* result in increased sales and hence in increased profits. But [this] ignores the restraining effect upon A & P's retail competition during the interval required for increased sales to reduce the expense rate.[9]

In what sense is this selling below cost? Suppose that consumers are so quick to see, compare and decide, that A & P receives the increased business immediately, and overnight moves from one profitable position to an even more profitable one. In that case, there is never any loss and never any selling below cost. But of course this may not happen. Consumers may or may not respond; costs may fall with higher output or they may rise. As the diagram shows, the instant result of a price reduction may be a (proportion-

7. *Brief for the United States* (hereafter cited as *GB* for "Government Brief"), pp. 110 and 837; our emphasis. See also *ibid.*, p. 1088.

8. U. S. Circuit Court of Appeals, October term 1947, No. 9221, *GB*, p. 373. Italics in original.

9. *Ibid.*, p. 384. Italics in original.

ately) much larger profit reduction or an actual loss.  A businessman trained in economics has well said that "it takes courage as well as sagacity to gamble on a rate of demand that will bring forth the volume needed for lower costs."[1]   Competitive price cuts of any magnitude typically (1) involve the risk or expectation of "selling below cost" until and unless the market responds; and (2) aim at larger profits (or smaller losses) through larger volume.  If A & P's price cutting is illegal, then most price competition, where profit margins are narrow, is illegal.

The A & P Company, in order to serve its own interests, absorbed progressively fewer resources per dollar or per ton of the goods it handled, and passed along to consumers a large part of the savings.  Obviously A & P sold "below the cost" which it *might* have incurred if it had preferred to stay at the old inefficient level.  The Government insists that it *should* have stayed there.  An "honest retailer" would attempt to "price his merchandise in the traditional American way, that is, cost, plus expenses, plus a profit."[2]  Thus cost at some predetermined level of output is taken as a basis to be maintained without change under penalty of law, and price is to be set by adding conventional markups.  Economists have been discussing "full-cost" pricing for at least a decade, and they are far from agreement on how widespread a practice it is.  But all parties have considered it as a departure from competition, and a restraint upon the operation of the market.  Yet we now witness its establishment as the standard for competition!

There are two interesting collateral issues.  (1) Suppose A & P had tried to attract business not by cutting prices but by making advertising and promotional expenditures which reduced net revenues by the same amount.  They would then have incurred the same losses until the campaign proved successful — if it ever did.  Would this be illegal?  If so, then almost any method of attracting business which costs or may cost money would be illegal.  But if it were legal, then it would be permissible for a firm to seek larger sales by enriching advertising media, and possibly misleading or confusing consumers; yet it would be illegal to seek business by giving consumers more goods for less money.  Further comment would be superfluous.

(2) Even if the A & P sales policy is nothing but price competition, it is legitimate to ask whether that competition might drive

1. Oswald W. Knauth, "Monopoly Reconsidered," *Political Science Quarterly*, December 1945, p. 566.
2. *GB*, p. 405.

all or most other firms out of business and allow the one or few survivors to exploit both growers and consumers. In other words, *if* there are no limits, or very wide ones, to economies of scale, competition becomes self-destructive, for the larger a firm, the lower become its costs; the lower its costs, the more it can cut price to gain more business and further lower costs . . . and so on. *But unless the assumption of no limits is true in fact*, there is no justification for asserting the consequences. Unless there are concrete facts to show that monopoly or oligopoly will certainly or at least probably result, the Department of Justice is in effect attacking competition on no better ground than an economic generalization which may be quite inapplicable to the case. It was a great lawyer, Mr. Justice Holmes, who remarked that "general propositions do not decide concrete cases."

All A & P stores in the United States and Canada are grouped by the Company into local "units" which coincide fairly well with metropolitan districts and with local market areas. The Government insists, rightly in my opinion, that it is the share of these *local* markets rather than of the national market which is relevant to any consideration of monopoly power.[3] If, now, for each year, that particular "unit" is chosen in which A & P stores did the highest percentage of the food business, we have a time series showing the upper limit of A & P market control for the years under review. In 1925 the highest unit was 17.14 per cent; successive peaks were, in 1935, 19.52 per cent; 1938, 20.51 per cent, and the third quarter of 1941, 19.81 per cent. The Company's share of the national market has at best remained stable since the early 1930's.[4] Yet the statement is made that "the picture . . . is one of an ever broadening and ever ascending spiral of monopoly and trade restraint in the hands of A & P."[5] It is highly significant that in a brief of 1100 pages there is not even one table presenting either the dollar or physical volume of sales or the share of the market, for the years in question, in the several units as described above.

But all that is in the past. What is indicated for the future by the structure and cost conditions of the industry? Are there such great economies of scale not offset by difficulties of co-ordination? If prices are raised by the surviving monopolists, will they be able to keep out new entrants, except by the expensive method of keeping prices down everywhere? The Anti-trust Division did not attempt

3. *GB*, pp. 1076, 1093–4.
4. *DF*, Appendix, Tables 2-A, 2-B.
5. *GB*, p. 1076.

to answer such questions; and they made strenuous and largely successful attempts to keep anyone else from asking them. Expert testimony that the economics of food-retailing rendered monopoly impossible was protested and often barred from the record not because it was false or misleading but because it was — of all things — *irrelevant!*

## II. INTEGRATION

The integration of the A & P Company has also been used to support the charge of "selling below cost." The great bulk of the Company's profit has been derived from manufacturing, the collection of advertising allowances and quantity discounts, operations of buying offices, and the like.[6]  The alleged restraint of competition is best put in the words of the Government brief:

> ". . . The profits from the non-retail end of the business . . . subsidized the retail business, so that the latter could operate at an uneconomic profit rate, *a privilege not possible to A & P's competitors.*  This, the Government contends, is an inherent abuse of the vertical integration of A & P's System."[7]

### *The Meaning of Intra-Firm Profits*

Certain items are excluded from retail profit for purely formal accounting reasons and must be added back. Advertising allowances and quantity discounts are reductions in the cost of goods purchased, and to call purchasing a non-retail operation is absurd.[8]  Thus we

6. The Department of Justice attempted, quite unnecessarily, to make profits derived from retailing appear even smaller than they were. Their statistical procedures can only be called deplorable. See *DF*, pp. 397–401.

7. *GB*, p. 86 (emphasis ours).  Compare also the following: ". . . Profits from all operations of the System are siphoned into its retail stores in order to offset uneconomic retail profit rates. The effect on competition of this unreasonable subsidization of retail operations is apparent in the relationship of profits from retail sales to the subsidies from non-retail sources" (p. 87). "As non-retail profits increased, Headquarters reduced the amount of profits from retail operations in order to place its stores in a position of price supremacy over competitors. By 1942, the crediting of non-retail profits to retail operations enabled A & P to operate its stores with inconsequential profits on retail sales" (p. 94).  ". . . The profits from A & P's non-retail operations are credited to their retail operations. . . . These direct subsidies make it difficult, if not impossible, for less strongly integrated, or for non-integrated competitors engaged only in retailing, to compete or survive" (p. 1075).

8. The Government alleges that A & P was able to coerce suppliers into discriminating in its favor. (This will be discussed in the next section.) The advertising allowances and quantity discounts contain these discriminations. To charge both that A & P burdened competition by receiving special concessions and that it used these concessions to subsidize its retailing and to burden competition is simply to say the same thing twice.

Incidentally, if we add the allowances, discounts, and buying-office profits back into the profits of local A & P areas — and it is pure, misleading formalism

are left largely with the substantial profits attributed to A & P manufacturing. That is, goods have been transferred from the manufacturing to the selling department at market prices, and profits or losses calculated for the A & P factories. The purpose of the Company's accounting system has undoubtedly been to picture its manufacturing and retailing separately, as if each one sold or bought in the open market rather than to and from each other.

But in fact the manufacturing plants do not sell, and the stores' warehouses do not buy. To pretend that they do is an accounting fiction — a useful and proper one but a fiction nonetheless. Now the law has numerous fictions of its own. A group of people is treated as a single "body" — this is the literal meaning of "incorporation." Again, all men are presumed to have an encyclopedic knowledge of the law, and ignorance is therefore no excuse. In economics, there are such fictions as perfect competition, the stationary state, or the instantaneous multiplier. They all have their place as aids to logical thinking.[9] But to treat them as realities is a caricature of logic. There is as much point in speaking of the profits of the A & P plants as in discussing the temperature or blood pressure of a corporation. No man can make a profit by selling to himself.

All A & P operations, up to the point where goods are alienated in return for hard cash, are designed to reduce the cost of these goods. Integration, if successful, is one such method. But the mere existence of *some* accounting profit is no indication of success. For example, in the middle 1920's A & P was on the point of giving up some of its profitable manufacturing facilities because the capital employed there could apparently draw a higher return in retailing. Had this continued to be the case, A & P would have been worse off than a retailer who had devoted an equal capital exclusively to retailing. Their "subsidized" retail stores would have been at a disadvantage in competing with the "unsubsidized."

The Department of Justice proclaims again and again that it is not opposed to integration as such, only to its misuse. Its sincerity is not questioned. But it simply has not realized that any integrated company whatever — in steel, automobiles, meat-packing, textiles,

to do anything else — then most of the bases for thinking that the Company sold below cost in selected geographical areas in order to drive out competition apparently disappear. But there is general agreement that such conduct, if it *did* exist, would be an unmitigated burden on competition. Since no issue of principle is involved, this charge has not been discussed at length.

9. Cf. Morris R. Cohen, *A Preface to Logic* (New York, Henry Holt & Co., 1944), ch. 5.

distribution — could keep its books in the same way as A & P simply by setting a market price on goods transferred from one department to another;[1] and that, so long as the earlier stages of any integrated company could be said to earn anything, however little, these fictitious "earnings" could be "siphoned" to abuse competition in the later states. And this would be the case even when possession of the earlier stages was a burden on the integrated company because they were earning too small a rate of return!

The only way to reconcile the legality of integration with the illegality of "subsidizing" — for they are simply two names for the same thing — would be to throw the whole burden of blame on the accounting system. Either any and all integration is illegal, or else a rational form of accounting is illegal, or both. This is the only possible implication of the Government case.

*Integration and Public Policy*

Thus far we have been concerned only to demonstrate that integration by any name is the same thing. But the issue is larger than mere confusion in the Department of Justice. Integration *can* be objectionable. A business concern may continue for years to make things which could be made more cheaply outside. In this case, if its books were set up like A & P's (and if practicable they certainly should be), there would be little or no profit "earned" at the earlier stages. The Department of Justice would be compelled to approve, since there would be no earnings to "siphon" for "subsidizing" later states. To an economist, on the other hand, it would seem to be a clear case of wasted resources and of competition failing to operate.

So much for abstract possibilities. Economic policy should be based on actual facts about the relative efficiency of integrated and non-integrated firms. The Department of Justice made no attempt to find them. But relative profitability may perhaps be used as a first approximation to relative efficiency. If we use some Federal Trade Commission data, and make the usual allowances for errors and discrepancies, it is my impression that A & P factories have been so much more profitable than large independent manufacturers of food products (who are themselves more profitable than the

1. The doubtful reader is urged to consult D. H. Wallace, *Market Control in the Aluminum Industry* (Cambridge, Harvard University Press, 1937), ch. VIII, where the technological advantages of integration are analyzed. Every such advantage is expressed for the business firm in costs lower than suppliers' prices. Hence there arises a "profit" on making the goods at home.

industry as a whole), that they are simply in a different class.[2] For this there seem to be two obvious reasons. First, the manufacturing plants have an assured market and can run along steadily at something like the optimum level of operations. There is no uncertainty about the market and no sudden and expensive adaptation to it. Second, and in part implied already, there are none of the large selling and advertising costs of the non-integrated food processors. Geoffrey Shepherd has referred to the "stupendous" cost of transferring ownership from one stage of distribution to the next.[3] A & P has completely eliminated them from its own organization. This saving benefits both A & P and the economy as a whole.[4] These are the gains which competition is supposed to bring about. The law evidently regards them in a different light.

It might be said that the cost of maintaining chronically idle capacity was a social burden that *someone* had to meet; that the Company was eliminating only part of this cost and shifting the rest onto its suppliers. But this also implies that A & P turned to these suppliers when demand and prices were high and delivery and quality relatively poor, and did not buy when it was cheapest to do so.[5] In other words, it is argued that the suppliers have become "specialists" in the bearing of risks in demand, and must be compensated accordingly — otherwise, if prices fall too low in slack times, some of them may be forced out of business, thus causing scarcity and unduly high prices during revival, an excessive influx of new capacity, and so on around the wasteful cycle. Does this actually happen in the food trade? It would be highly relevant to know as an element in judging the industry. But there is no evidence to suggest that the Department of Justice has ever considered the question.

### III. BUYING POLICY

The buying policy of the A & P Company may be the most important single issue and is undoubtedly the most complex. It is necessary to disentangle two distinct aspects: (1) coverage of the market, and (2) inducing sellers to discriminate.

#### 1. *Coverage of the Market*

(a) *"Demoralization."* An imperfect market, containing not a single price but a spread of prices, can be exploited by a large buyer

2. See *DF*, pp. 312–313 and 407–410.
3. Geoffrey Shepherd, *Marketing Farm Products* (Ames, Iowa State College Press, 1946), p. 391.
4. See footnote 2, above.
5. I am indebted to Sidney S. Alexander for pointing this out to me.

who can be "all over the place" at any given moment and take the better offers. The effect of such a buyer is to narrow the range of prices, lower their average, and make for a single price in the market.

Evidence both in and out of the record indicates that sellers have a dislike for such buying offices, a dislike which is out of proportion to the size of their operations.[6] Letting the big buyer inside the market is considered very bad because concessions granted to him have such an unfortunate tendency to spread. Sellers who might otherwise receive higher prices because of their customers' ignorance of lower prices elsewhere in the market are forced to meet these lower prices. The large buyer may undercut them directly by reselling some bargain offers which he cannot use. Or he may set off vigorous price competition at a later stage in the productive or distributive process. A & P did both. It is not difficult to understand sellers' and competitors' resentment at such "demoralization." But it is difficult to see why anyone concerned with the general welfare should share it.

(b) *Brokerage.* Food processors and packers commonly sell to distributors through brokers who are paid a certain percentage of the list price. A & P has attempted to act as its own broker and to collect its fee in the form of an equivalent reduction in the list price, leaving the net price to the seller unchanged. The Government has devoted volumes of testimony and exhibits simply to demonstrating the fact that a brokerage equivalent was allowed to A & P buyers, and especially to its produce-buying subsidiary. Very rarely are these allowances said to be excessive and to constitute price discrimination, which will be discussed later. Assuming, as we do throughout for the sake of a policy discussion, that the facts are as alleged, why is it wrong for a firm to do its own buying and pay a lower gross price?

Consider two buyers at the same location. One of them buys goods delivered, the second buys f.o.b. and performs the transport function himself. If both of them pay the same price to the seller, it is obvious that the second is suffering discrimination. He is paying for a service he himself performs, and the seller is getting a higher net return from him than from the other.

The middleman or broker function is closely analogous to the transport function. If A & P paid the same prices as those who

6. See *DF*, pp. 126, 347, 484–485, and Appendices 5 and 7; also Charles F. Roos, *NRA Economic Planning* (Bloomington, Principia Press, 1937), pp. 277–279, 293–304, 408, 503, 505; Charles Albert Pearce, *NRA Trade Practice Programs* (New York, Columbia University Press, 1938), p. 78.

bought through brokers, the net return to sellers would be greater on sales to A & P.  In many cases, strong sellers, strengthened further by the Robinson-Patman Act, forced them to buy at list price and pocketed the difference: correspondence with Del Monte[7] makes it beautifully clear that the packer was collecting phantom brokerage in precisely the same way that steel or cement mills used to collect phantom freight — for services never rendered.  This is what the Department of Justice is trying to bring about.  The only possible result of its policy is to enforce systematic patterns of price discrimination which weaken competitive pressures on sellers.  It is a perfect analogy to zone pricing systems, with identical prices at different locations.  But we need no analogies.  The policy of implicit discrimination is writ large in the Sugar Institute code which the Supreme Court struck down in 1936.[8]  And this is mentioned as only the most notorious among many whose common purpose was to preserve an "orderly" or "stable" price structure from the pressure of buyers who might force prices down and even — worse thought — set off price competition at the reselling level.  The Department of Justice, in this case, is in full sympathy with the stabilizers, even as, in other cases, it prosecuted them.

Just as a basing-point or zone system weakens the urge of buyers to find cheaper transportation methods, so a system of identical list prices to all buyers (which means diverse net prices) weakens the urge to find or set up cheaper distribution methods.  But can A & P buying offices really do a better job than independent brokers?  As usual, the Department of Justice does not ask the question.  It is possible, however, to compare the income statement of A & P's produce-buying subsidiary with OPA data for independent wholesalers.  The A & P subsidiary handles a longer flow — from the field to the warehouse, as against the independents who operate at terminal markets — yet its gross margin, the price of its service, is actually *one-third to one-half as large* as that of the firms who do a much smaller job.  The superior efficiency, in fact, is hardly in dispute.[9]

Any *net* advantage to the Company from operating the buying offices would of course be in their *net* profit.  As we have seen, the Department of Justice appears completely to ignore this elementary fact.  For if they were aware of it, why the heavy emphasis on

7. In the record as Government Exhibit 1829-A.
8. See *DF*, p. 292.
9. *DF*, pp. 328–330; the most nearly comparable group are "full service wholesalers."  See also pp. 336–337.

gross brokerage payments? *If* we were to assume that A & P buying offices needed no personnel, premises, or equipment, and cost the Company nothing, *then* there might be sense in speaking of brokerage payments in the gross as a "rebate" or "a preferential buying advantage."[1] But only under such an assumption would the conclusion be warranted.

(c) *Profits of Buying Offices on Re-Sale.* But there is a surprise in store for us. After completely neglecting the expenses of buying operations in order to make the advantage look perhaps fifty times as great,[2] the Government turns around to say that competition was *also* abused over and above this advantage by the net profits of the buying offices. About three-fourths of these "profits" were on re-sales to the A & P and are therefore as purely formal and meaningless as the manufacturing "profits" discussed earlier.[3] But the Company also profited by re-selling goods to competitors, thus in effect drawing a "subsidy" from the very firms with whom they were competing.

The confusion is fairly obvious. Had A & P withdrawn its personnel and investment from buying and devoted them instead to some other profitable occupation, competitors would have had to pay no less for produce, and probably more, while A & P would have been just as much strengthened against them. For example, the very aggressive persons in charge of produce-buying might have been assigned to expand the retail activities of the Company — say, to opening new supermarkets after the war when building materials again became plentiful. Their profits would have "subsidized" the other stores just as effectively. But why stop there? Under penalty of contradicting itself, the Department of Justice must insist that at any moment of time half the stores are "subsidizing" the other half; or rather that every individual store is "subsidizing" every other. Nor need we confine ourselves to the food industry, nor to big business. For example, the repair department of an automobile

1. *GB*, p. 225.
2. *DF*, p. 330.
3. On certain transactions, produce-shippers did not sell directly to A & P but did consign their wares to its produce-buying subsidiary to be sold at the best price obtainable. Some portion of these goods might be and doubtless was eventually sold to the retailing organization. Within limits, there would be a presumption that such shippers would run the risk of receiving a price somewhat less than the market. The obvious remedy would seem to be some kind of compensation device. In fact the produce-buying subsidiary, doubtless under pressure, waived the collection of brokerage on such sales. They were all the more ready to do so because such sales could not have amounted to more than about 5 per cent of their transactions.

dealer gives rise to profits which "subsidize" and give an unfair advantage to his sales department in its rivalry with dealers who have no repair shops.

If even such tiny horizontal integrations are in restraint of competition, consider what great vertical encumbrances must be swept away. In the Aluminum case, the Government, at the cost of immense labor and time, was able to prove that Alcoa had abused competing aluminum fabricators by charging them a very high price for ingot and then undercutting them on finished products. But under the new doctrine it would have been necessary to prove only that Alcoa had made some profit, large or small, on sales to fabricators. This profit was a "subsidy" in precisely the same way as the profits of A & P buying offices. Hence it would suffice as proof of a burden on competition.

To sum up the new doctrine: it asserts illegal restraint of competition to exist in any business firm which (1) can be set up as more than one accounting and profit-making unit, *or* (2) directly or indirectly sells anything to a competitor. We shall seek a long time, I fear, before finding any firm which is not breaking the law. This suggests a synthesis of the two criteria. They are both logical consequences of the notion, profoundly anti-capitalistic in spirit, that business profit is somehow a subsidy or gift; it follows that the return on every separate transaction is a subsidy to every other, making each one a separate act in restraint of trade.

## 2. *Inducing Sellers to Discriminate*

(a) *The Pattern.* The effects of a pattern of price discrimination upon more and less favored buyers and upon competition cannot usually be measured; even the qualitative effects cannot often be established by direct evidence. Hence the Government demonstrates their existence by deductive reasoning. The major premise is that if sales are made to one buyer at a price lower than that charged others, price to the others must necessarily be further increased in order to offset the loss or the smaller profit on the sale to the favored buyer. Or, what comes to the same thing, price concessions to some buyers can never lower the average gross revenue of any seller. This appears to hold at all times and under all circumstances with no qualifications whatever.[4]

4. "A & P not only benefits by the discriminatory purchase price, but the supplier giving the preference must charge A & P's retail competitor a price sufficiently higher than would otherwise be the case to absorb the loss or reduction in profit incident to his sales to A & P." *GB*, p. 1075.

"The price differential between A & P and its retail competitors was

As I have written elsewhere:

"It would be a strange business man who was able to raise prices to Buyer X, but waited until he lowered them to Y. And even aside from this curious behavior: it would take no elaborate poll of business men to establish that a price reduction may come out of profits. Furthermore, there is the possibility of a very common occurrence: of a loss of revenue inducing a searching of heart and of wastes, and leading to lower costs. That, of course, is the way of competition. The theory of 'necessary recoupment' assumes no excess capacity to exist, and no profits, above the bare minimum needed to keep the firm in operation. This amounts to assuming the fictitious perfect competition — which is logically incompatible with price discrimination anyway!"[5]

Is it not clear that the Government has formulated a highly abstract theory which could be correct only on narrow assumptions rigidly adhered to (and not even then); has constructed on the basis of that theory a drastically simplified economic model (which also happens to contradict itself), and has then in all innocence applied the model directly to a complicated reality? Once admit that competition is not perfect and that most firms are usually working with some excess capacity and under less than optimum conditions; once admit what the Department of Justice seems to find it impossible to conceive, namely that output and cost are not fixed, and it is easy to see how these latter can be altered so as to benefit, though doubtless unequally, every party to every transaction. Given its aggressive selling policy, discrimination in favor of A & P has, in fact, probably resulted in higher output produced at lower cost and sold at lower prices to consumers.

  (b) *The Market Process.* This is the more interesting and important aspect. In imperfect markets, such as those in which A & P buys, an integral part of the process of price reduction is unsystematic, buyer-enforced price discrimination. To forbid this kind of discrimination is simply to subvert competition. In the A & P record itself the emphasis on secrecy is sufficient evidence of sellers' well-justified fear of general reductions if the news gets out.[6] Looking more widely, the long history of many trade associations, their opposition to price discrimination, and their unwearying attempts to stamp it out,

enhanced further because of the necessity that suppliers increase the price to competitors over and above the price differential in order to recoup losses incurred in sales to A & P." *Ibid.*, p. 1077.

  See also *DF*, pp. 478–479.

  5. "Effective Competition and the Antitrust Laws," *Harvard Law Review*, September 1948, p. 1331. See *DF*, pp. 479–482 for a demonstration that the Government theory assumes (1) pure competition (2) a stationary industry, (3) identical cost and demand conditions for all firms in it and (4) identical sources of supply for all buyers.

  6. *DF* pp. 214, 282–283.

should speak for itself. So should a remarkable pamphlet issued by the National Association of Manufacturers in its less cautious days.[7]

The Government objects to this kind of discrimination, but as already seen,[8] it blesses — indeed, it requires — precisely those systematic, ordered patterns of discrimination which collusive oligopolies have traditionally used to keep price-cutting out in the open and therefore at bay.

(c) *Advertising Allowances*. Concessions to A & P frequently were in the form of payments or allowances in return for including and featuring the supplier's product in A & P advertising. A portion, but by no means all, of these allowances was simply re-expended in the purchase of advertising space. This implies one or both of two things: (1) A & P took payment for advertising that would have been done anyway and which required no additional (marginal) expense, or else (2) it was able to buy advertising space at much cheaper rates than its suppliers and re-sold it to them at a profit. In either case the effect was to reduce the supplier's advertising bill, and reduce, by a somewhat smaller amount, the net price which the Company paid for its goods.

(1) In the first case, advertising which was an overhead expense to the Company (since A & P advertising nearly always had to feature some specific commodities) could also be made to serve, and did serve, its suppliers. Had the Company sublet parking space in their garages to their suppliers' trucks, and been paid a fee therefor, the economic effect would have been no different from an advertising allowance (or, of course, a brokerage payment). In either case an existing facility would be exploited jointly by more than one party; all would share in the expense, and all would be better off because they were able to save space — advertising space or garage space. In short, joint advertising was a more economical way of doing business, and as such it made things uncomfortable for competitors. If the lion's share of the economy passed to A & P, then in view of its aggressively competitive selling policy most of it passed on to consumers.

(2) Where A & P was able to buy advertising cheap and sell it dear, the cost to suppliers of such advertising was less than by alternative methods; otherwise they would not have bought it. Again, both parties could benefit by reducing their combined advertising bill and sharing the saving.

But in either case, the seller, instead of buying more advertising,

7. See the *Harvard Law Review*, *loc. cit.*, for a summary of the evidence.
8. Above, p. 249.

in effect cut his net price for the sake of A & P patronage. To him, a cut in price and an advertising expenditure stood on the same footing — each was a temporary sacrifice for the sake of a future gain, and the choice between them depended on which promised a greater return for a smaller sacrifice. But from the social point of view, price reduction is a direct benefit of competition, whereas advertising expenditure is part of its cost.

Thus the collection of "excessive" allowances meant lower prices to A & P, and to the consuming public; and it encouraged or forced other retailers to seek and pass along the same economies. But to the Government, A & P was forcing its suppliers to pay for its advertising. It might equally well claim that by exacting lower prices in some other form, A & P was forcing its suppliers to meet its payroll or its bill for electric power. But doubtless many competitors and suppliers of A & P would agree with the Department of Justice that there was something "foul, strange, and unnatural" in cutting prices to increase sales.

(d) *The Leverage of Private Brands.* The buying, selling, and manufacturing policies of A & P are all epitomized in one aspect which the Government itself heavily emphasizes. The Company has always the alternative of (a) buying branded goods, (b) putting its own labels on unbranded items or (c) manufacturing for itself, again for its own labels. The consumer who buys under its labels is no longer restricted to buying food plus soap operas plus the expensive services of a large selling force, all in one package; he has the additional alternative of buying food alone and passing up the associated pleasures if he so desires. The competition from private labels has tended to lower the prices of branded goods and to inhibit retail price fixing by agreement between manufacturer and retailer.[9]

In general, this seems like a very desirable state of affairs. But let us consider the concrete case to which the Government gives most attention. The profits of the Ralston-Purina Company on sales to A & P were very large. There were only two other firms who manufactured corn flakes for private label, and it is easy to imagine that buyers may be over a barrel when there are only three sellers in the market. But A & P drew Ralston's attention to the fourth alternative of manufacturing for itself, and consequently received a cheaper price. As a result of this reduction, prices of advertised corn flakes were also reduced.[1] To me, this is a clear

---

9. Federal Trade Commission, *Report on Resale Price Maintenance* (1945), pp. 282–286, 297, 298, 636.

1. See *DF*, pp. 273–277, and p. 205.

case of workable competition benefiting the consumer, both by affording him a wider choice and by operating to cut prices all around.[2] According to the Government, it is coercion, a threat to boycott (sic), the exercise of monopoly power, and should be extirpated.

## IV. CONCLUSIONS

In the A & P case, the Government has reached further into the details of business policy than in any other case with which I am familiar. By way of summary, let us ask: What kind of behavior is it trying to enforce? What theory justifies its prescriptions?

*1. Behavior*

(a) *A completely rigid cost-price position.* A firm should not cut prices in the hope of attracting more business. An "honest retailer" would attempt to "price his merchandise in the traditional American way, that is, cost, plus expenses, plus a profit."[3] And cost is, and must continue to be, fixed at what it happened to be yesterday. One must not cut price below yesterday's cost for the sake of more business at a remunerative lower cost tomorrow.

(b) *Rigidity also of economic function.* A retailer ought not to wholesale or to manufacture; for efficient operation in these fields would give him an advantage over his rivals which it is "unfair" to seek. But there is no objection to his doing an inefficient job and gaining no advantage.

(c) *A ban on the promotion of price competition.* To spread lower prices throughout the market, or to obtain lower prices, or to induce a supplier to spend less on advertising and pass the savings along to the consumer[4] — all these practices which disturb the placid existence of men staking out little areas of the market as their private preserves, are disapproved.

(d) *A ban on discrimination of the sort which may and usually does promote competition.* However, systematic discrimination designed to

2. "Thus it becomes peculiarly important to preserve an open field for the competition of different kinds of distributive outlet, not only because the distributive trades need competition but because such diversities protect competition in manufacturing." Corwin D. Edwards, "Can the Antitrust Laws Preserve Competition?" in *Readings in the Social Control of Industry* (Philadelphia: Blakiston Co., 1942), p. 21.

Some confusion may be avoided by pointing out that Mr. Edwards has not been associated with the Department of Justice for the past seven years. At present he is chief economist for the Federal Trade Commission.

3. Cited above, p. 242.

4. I.e., when a supposed advertising allowance is really a disguised price reduction.

suppress more efficient distributive methods and the more active price competition which they set off — this kind of discrimination is not only approved but required.

This cost-plus philosophy is not exactly novel. It is the usual program of collusive understandings to mitigate the rigors of competition, and it flourished like the green bay tree under the NRA. Now it emanates from the Department of Justice. The Devil finds refuge in the most unexpected places!

## *2. Theory*

All this is in the name of fighting a nebulous bogey called "buying power." The greatest strength of the Government's economic theory is its unbelievable vagueness. When the weary investigator finally untangles the doctrine, its fallacies are trite and obvious. That it flouts the canons of academic economists is less important than that it is offensive to common sense. I doubt whether many ordinary citizens can swallow the notion of chain stores as monopolies against the public interest. There is a real danger that such legal actions will discredit the wider interpretation of the antitrust laws and invite crippling amendments.

Why has the Department of Justice done so badly? In a sense, the answer is obvious: they have confused the maintenance of competition with the protection of particular competitors. But the two are usually (though not always) antithetical: no strong competitor, however pure and upright his methods (no claims are made for A & P), however beneficent his achievements, could avoid affecting adversely other competitors and even putting some of them out of business. It is therefore no criticism of competition that these things have happened. In general, it may be said that most or all of the economic fallacies in this case are of a kind which should be quickly obvious to any trained and competent economist. Intelligence alone is not sufficient for straight thinking on economic issues — there is needed also an apparatus of economic concepts and principles, in other words, a theory. The very little that economic theory can say about the real world as yet may, as this one case proves, be very important, and be neglected at one's peril. Doubtless the economic profession would be more respectable if it could make more positive contributions, and supply more ready yardsticks of workable competition. But there is, in addition to a "supply" problem, a "merchandising" problem. It is difficult to sell yardsticks to people who are trying to buy clubs.

The state of mind of a government enforcement bureau is under-

standable. A feeling of weakness and of tiny size in relation to huge tasks, and a realization of how few cases can be tried, puts a premium on swinging at any target which seems likely to go down, preferably with a loud noise which may impress congressional committees. If this article's thesis is correct, the remedy lies not in smaller but in larger appropriations; but money is clearly not enough. To enlarge and strengthen the antitrust agencies and expect them to become more reasonable because more effective — and vice versa — does not seem too forlorn a hope; but we must exact as a price a program which makes consistent sense.[5] At present, if this case is indicative, the Antitrust Division is not attacking objectionable or excessive market control, and it is not seeking workable competition. It not only lacks standards for both; its actions appear not even to recognize that either phenomenon exists.

The economist is as much to blame in this situation as is anyone else. The A & P case is at once too simple and much too sophisticated. I should like to, but cannot, disagree with two professors of law who considered that only the intellectual climate generated these past fifteen years by economists could have permitted others to see illegal monopoly in such cases. We let loose the notion, perfectly correct in itself, that every market contains some elements of control. Under a simple — very simple — construction of the Sherman Act, therefore, it becomes possible to view every market as tinged with illegal monopoly. The other side of the shield is evidently that elements of monopoly are no bar to effective and socially desirable competition — they may even be inextricably bound up with it. It behooves economists to get on with the job of making more explicit the standards of public policy in this borderline area.

<div align="right">M. A. ADELMAN.</div>

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

5. Milton Katz urged this consideration very strongly upon me before I started intensive work on the case. I now remember with no little chagrin that I was not greatly impressed.

# THE MULTIPLIER IN A TRI-FISCAL ECONOMY

## SUMMARY

Introduction: interdependence of federal, state, and local revenues and expenditures, 258. — Definitions, 259. — Assumptions, 261. — Multipliers, 263. — Interpretations, 266. — Applications, 268. — Examples, 270. — Conclusions, 272.

The task of estimating the effects of governmental expenditures on national income in a tri-fiscal system such as that of the United States — with its federal, state and local fiscal authorities — is complicated by the great dependence of state and local expenditures on the national income. Such dependence is probably not a direct one but operates through changes in tax revenues. State and local tax revenues as a whole have a significant income-elasticity. As for state and local expenditures, they are influenced by tax revenues which in turn are controlled by national income which is affected by all governmental expenditures — federal, state and local. The behavior of state and local tax revenues in relation to national income can probably be used, therefore, to estimate the marginal propensities of state and local governments to spend. These marginal propensities may then be employed in income-estimation work along with the usual array of marginal propensities, such as those concerned with consumption, investment, foreign trade, etc.

The marginal propensities of state and local governments to spend may be assumed to be positive — induced state and local expenditures to rise as national income rises and to fall as national income falls. The marginal propensity of the federal government to spend may, however, be assumed to be negative — induced federal expenditures to fall as national income rises and to rise as income falls. The sum of the state and local marginal propensities to spend may more than offset the federal marginal propensity to spend.

This emphasizes the necessity of breaking down the governmental element in the multiplier formula into federal, state and local components. The federal government may reduce its rate of spending as income rises, but state and local governments may behave in precisely the opposite way. It is well known that state and local governments may nullify counter-inflationary or counter-deflationary fiscal efforts of the federal government. It is also important to take account of state and local repercussions in order to economize the federal expenditures and deficit required to achieve a given increase

in income when a counter-deflationary policy is being pursued. The induced increase in state and local expenditures makes it possible to achieve a desired increase in income with smaller federal expenditures than would otherwise be possible. Similarly, a counter-inflationary policy can achieve a desired reduction in national income through a smaller reduction in federal expenditures and deficit (or a smaller increase in surplus) than would otherwise be necessary. The induced reduction in state and local expenditures reinforces the counter-inflationary effects of the reduction in federal expenditures.

The multiplier analysis developed here is designed to provide an analytical framework which permits explicit formulation of these effects and their quantitative measurement. There are assumed to be three fiscal authorities, A, B and C, each of which has the power to tax, spend and borrow. They are all members of the same national unit just as the federal, state and local fiscs are part of the governmental system of the United States.

## DEFINITIONS

The following symbols and terms are used in the analysis. Certain elementary assumptions directly implied in the definitions are included in this section (equations (13) — (19) ).

Let $Y$ = change in money national income

$R_a$ = change in A's tax revenues

$R_b$ = change in B's tax revenues

$R_c$ = change in C's tax revenues

$R$ = change in total tax revenues, i.e.

(1) $$R = R_a + R_b + R_c$$

$E_a$ = change in A's autonomous or income-inelastic expenditures, i.e. those not influenced by national income changes (e.g. national defense expenditures or interest charges).

$E_b$ = change in B's autonomous or income-inelastic expenditures

$E_c$ = change in C's autonomous or income-inelastic expenditures

$N_a$ = change in A's income-induced or income-elastic expenditures, i.e. those influenced by national income changes (e.g. relief payments or operating costs)

$N_b$ = change in B's income-induced or income-elastic expenditures

$N_c$ = change in C's income-induced or income-elastic expenditures

$X_a$ = change in A's total expenditures, i.e.

(2)
$$X_a = E_a + N_a$$

$X_b$ = change in B's total expenditures, i.e.

(3)
$$X_b = E_b + N_b$$

$X_c$ = change in C's total expenditures, i.e.

(4)
$$X_c = E_c + N_c$$

$E$ = change in total autonomous or income-inelastic governmental expenditures, i.e.

(5)
$$E = E_a + E_b + E_c$$

$N$ = change in total income-induced or income-elastic governmental expenditures, i.e.

(6)
$$N = N_a + N_b + N_c$$

$X$ = change in total governmental expenditures, i.e.

(7)
$$X = X_a + X_b + X_c \qquad \text{and}$$

(8)
$$X = E + N$$

$D_a$ = change in A's budget deficit, i.e.

(9)
$$D_a = X_a - R_a$$

$D_b$ = change in B's budget deficit, i.e.

(10)
$$D_b = X_b - R_b$$

$D_c$ = change in C's budget deficit, i.e.

(11)
$$D_c = X_c - R_c$$

$D$ = change in total budget deficits, i.e.

(12)
$$D = D_a + D_b + D_c$$

$C$ = change in income-induced aggregate consumption in money terms

$\alpha$ = the marginal propensity to consume for the economy as a whole, i.e.

(13)
$$C = \alpha Y$$

Disposable income might be used instead of national income in this case. We would then have $C = \beta (Y - R)$ where $Y - R$ is disposable income and $\beta$ is the marginal propensity to consume out of disposable income. Since tax revenues are assumed to be linear functions of national income, we can assume in (14), (15, and (16), as we do in (13), that consumption is also a linear function of national income.

$\gamma_a$ = A's marginal tax productivity of national income (in multiplier language, the marginal "propensity" to pay taxes to A)

= the fraction of any change in national income which A receives in tax revenues, i.e.

(14)                     $R_a = \gamma_a Y$

$\gamma_b$ = B's marginal tax productivity of national income (in multiplier language, the marginal "propensity" to pay taxes to B)

= the fraction of any change in national income which B receives in tax revenues, i.e.

(15)                     $R_b = \gamma_b Y$

$\gamma_c$ = C's marginal tax productivity of national income (in multiplier language, the marginal "propensity" to pay taxes to C)

= the fraction of any change in national income which C receives in tax revenues, i.e.

(16)                     $R_c = \gamma_c Y$

$\lambda_a$ = A's marginal propensity to spend

= the fraction of any change in national income which A spends, i.e.

(17)                     $N_a = \lambda_a Y$

$\lambda_b$ = B's marginal propensity to spend

= the fraction of any change in national income which B spends, i.e.

(18)                     $N_b = \lambda_b Y$

$\lambda_c$ = C's marginal propensity to spend

= the fraction of any change in national income which C spends, i.e.

(19)                     $N_c = \lambda_c Y$

## ASSUMPTIONS

*Assumption I.* A may unbalance its budget with impunity, i.e.

(20)          $D_a = X_a - R_a \gtrless 0$   or   $X_a \gtrless R_a$

It will be recalled that (14) makes $R_a$ a function of $Y$. The present assumption means that $X_a$ is independent of $R_a$ and may be varied at will.

*Assumption II.* B and C do not exercise their power to borrow but always balance their budgets and always spend any money they receive, i.e.

(21)                $D_b = X_b - R_b = 0$    or    $X_b = R_b$

(22)                $D_c = X_c - R_c = 0$    or    $X_c = R_c$

Since (15) and (16) make $R_b$ and $R_c$ functions of $Y$, the present assumption means that $X_b$ and $X_c$ will equal $R_b$ and $R_c$ respectively and vary with $Y$ in the same way as do $R_b$ and $R_c$.

From equations (2), (14), (17), and (20) and putting $E_a = 0$,

(23)                $\gamma_a Y \gtreqless \lambda_a Y$    or    $\gamma_a \gtreqless \lambda_a$

From equations (3), (15), (18), and (21) and putting $E_b = 0$,

(24)                $\gamma_b Y = \lambda_b Y$    or    $\gamma_b = \lambda_b$

From equations (4), (16), (19), and (22) and putting $E_c = 0$,

(25)                $\gamma_c Y = \lambda_c Y$    or    $\gamma_c = \lambda_c$

It should be emphasized that assumption *II* is not required for the derivation of a "Multiplier in a Tri-Fiscal Economy." This assumption is made with a view to deriving the state and local marginal propensities to spend from the behavior of tax revenues. If this assumption is unrealistic it may be dropped and the marginal propensities to spend may be derived in some other way, perhaps directly.

*Assumption III.* A closed economy is assumed, and changes in private investment are assumed to be nil in the initial formulation. Changes in investment and exports are taken into account in equation (34).

*Assumption IV.* There are no independent or autonomous or exogenous changes in tax revenues. Any changes in tax revenues which take place are a result of the operation of the multiplier, i.e. are dependent on income changes induced by governmental expenditures.

It should be noted that $\gamma_a$, $\gamma_b$, and $\gamma_c$ take account of any and all tax revenues of A, B, and C respectively, which are variable with national income. The simple relationships of (14), (15), and (16), i.e. $R_a = \gamma_a Y$, $R_b = \gamma_b Y$, and $R_c = \gamma_c Y$, are assumed to describe adequately the dependence of the tax revenues on the national income. This points up the necessity of being certain that $\alpha$ is "pure" (net of all taxes, even sales and excise taxes whose revenues vary with national income). This is implicit in the multiplier analysis but must be emphasized here to avoid double counting, i.e. overlapping between $\alpha$ on the one hand and $\gamma_a$, $\gamma_b$, and $\gamma_c$ on the other.

*Assumption V.* There are no fiscal arrangements (e.g. shared taxes, grants-in-aid, tax credits, etc.) among A, B, and C    There is

some degree of fiscal interdependence, however, in that the tax revenues of each depend on the national income (under (14), (15), and (16) ) which in turn is affected by the expenditures of all under the usual multiplier relationship (28).  The introduction of formal fiscal arrangements alters the analysis in some respects and accentuates the desirability of taking account of state and local repercussions in income estimation.

*Assumption VI.*   Exogenous or autonomous changes in state and local expenditures are assumed to be absent, i.e.

(26)                          $E_b = 0$   or   $X_b = N_b$

(27)                          $E_c = 0$   or   $X_c = N_c$

This assumption can be removed easily, as in model *IIIc* and equation (33), but it is made here to facilitate emphasis on the multiplier effects of spending by the federal government and subsequent effects on state and local revenues and expenditures.

*Assumption VII.*   The respective marginal propensities are assumed to be constant and uniform for any level of expenditures. Modifications of this assumption may be made without serious impact on the theoretical argument.   Such modifications may, however, alter considerably the application in specific cases, especially if serious curvilinearity prevails and linear approximations therefore involve substantial error.

*Assumption VIII.*   The ultimate multiplier effects are considered here without regard to any specified calendar period.   Application of the results to a given calendar period would require modification to allow for relevant time lags.

### MULTIPLIERS

Various econometric models can be set up to derive multipliers showing the relation between autonomous changes in expenditures and resulting changes in national income with or without allowance for state-local repercussions or other induced changes.   In order to save space the equations comprising the models will be referred to by number wherever possible instead of being reproduced.

*Model Ia.*   Under assumption *III*, changes in national income consist of changes in consumption expenditures and governmental expenditures.   Neglecting any autonomous changes in consumption, we have

$$Y = C + X$$

Under (13),              $C = \alpha Y$      i.e.

$$Y = \alpha Y + X$$

Hence,

$$(28) \qquad Y = \frac{X}{1 - \alpha}$$

This is the usual multiplier formula whereby no separation is explicitly made of autonomous and induced governmental expenditures. In other words, it is implicitly assumed that all governmental expenditures are autonomous.

*Model I b.* If we do explicitly confine ourselves to autonomous governmental expenditures and assume that there are no changes in state and local expenditures, we have

$$X = E_a \quad \text{i.e.}$$
$$Y = C + E_a$$
$$= \alpha Y + E_a$$

Hence,

$$(29) \qquad Y = \frac{E_a}{1 - \alpha}$$

*Model II.* If, however, we allow for income-induced as well as autonomous federal expenditures, we have

$$Y = C + X_a$$
$$= C + E_a + N_a,$$

Under (13) and (17), we have

$$Y = \alpha Y + E_a + \lambda_a Y$$

Hence,

$$(30) \qquad Y = \frac{E_a}{1 - \alpha - \lambda_a}$$

*Model III a.* Where income-induced changes in state and local expenditures are also permitted, we have

$$Y = C + E_a + N_a + N_b + N_c.$$

By (13), (17), (18), and (19), this gives us

$$Y = \alpha Y + E_a + \lambda_a Y + \lambda_b Y + \lambda_c Y,$$

Hence,

$$(31) \qquad Y = \frac{E_a}{1 - \alpha - \lambda_a - \lambda_b - \lambda_c}$$

*Model III b.* Under assumption *II* changes in expenditures of state and local governments equal changes in tax revenues of those governments, such that equations (24) and (25) prevail, i.e.

$$\gamma_b = \lambda_b \text{ and } \gamma_c = \lambda_c.$$

Under assumption $I$, no such equality holds in the case of federal expenditures (except by accident). Hence (31) becomes

$$(32) \qquad Y = \frac{E_a}{1 - \alpha - \lambda_a - \gamma_b - \gamma_c}$$

*Model III c.* Removing assumption *VI*, all autonomous governmental expenditures — federal, state and local — may be taken into account, giving us $E$ instead of $E_a$, i.e.

$$Y = C + X.$$

From (8),

$$Y = C + E + N.$$

From (6),

$$Y = C + E + N_a + N_b + N_c.$$

From (13), (17), (18), and (19),

$$Y = \alpha Y + E + \lambda_a Y + \lambda_b Y + \lambda_c Y$$

Hence,

$$(33) \qquad Y = \frac{E}{1 - \alpha - \lambda_a - \lambda_b - \lambda_c}$$

*Model IV.* As a digression, we may temporarily suspend assumptions *III*, *IV*, and *VI* to obtain a more general multiplier formula.

Let $U$ = value of autonomous changes in all expenditures (both public and private, for domestic and export use) after allowance for any autonomous changes in tax revenues and imports in so far as these affect the magnitudes of the components of national income.

$I$  = change in income-induced private investment expenditures

$\eta$  = marginal propensity to invest, i.e. proportion of increased investment to increased income, i.e. $I = \eta Y$

$F$ = change in income-induced expenditures on goods for export. A change in income may induce such change in export expenditures directly (in some way) or indirectly by stimulating imports, thus increasing the ability of foreign countries to buy our exports. This is admittedly a shorthand way of dealing with a complicated subject like the foreign trade multiplier. The value of $F$ may, of course, be zero. The opening up of the economy in this way makes it necessary for us to redefine $C$, $I$, and $X$ so as to cover domestic expenditures on domestically produced goods only. If they covered domestic expenditures on all goods, $F$ would have to be redefined as exports minus imports.

$\mu$ = marginal propensity to spend on exports, i.e. proportion of increased exports to increased income, i.e. $F = \mu Y$.

It should be remembered that all autonomous changes in exports and imports which may affect national income are considered in $U$. Income-induced changes in imports are not considered because they are "leakages": any changes in such imports can affect national income only if they affect one of the propensities to spend.

$H$ = value of all income-induced changes in expenditures (both public and private, for domestic and export use), i.e.

$$H = C + N + I + F$$

Then
$$
\begin{aligned}
Y &= U + H \\
&= U + C + N + I + F \\
&= U + C + N_a + N_b + N_c + I + F \\
&= U + \alpha Y + \lambda_a Y + \lambda_b Y + \lambda_c Y + \eta Y + \mu Y
\end{aligned}
$$

Hence, a more general multiplier formula is

$$(34) \qquad Y = \frac{U}{1 - \alpha - \lambda_a - \lambda_b - \lambda_c - \eta - \mu}$$

Income-induced changes in private investment and in exports have been provided for in the usual way in the denominator of formula (34). In order to avoid obscuring the main point being made, such induced changes are not considered hereafter. The multiplicity of state and local governments in the United States might also be taken readily into account. We would then have a "multiplier in a multi-fiscal economy."

The temptation to simplify these formulae by substituting $\lambda$ for $\lambda_a + \lambda_b + \lambda_c$ should be resisted because that would obscure the part played by state-local repercussions in the multiplier process.

INTERPRETATIONS

We can see that equations (31)–(34) have some advantages in the theory of fiscal policy and in practical income-estimation work. For convenience, let us refer to A, B, and C as the federal, state and local governments, respectively. If we use the formula

$$(28) \qquad Y = \frac{X}{1 - \alpha}$$

we assume that the total of all governmental expenditures is given. This would be a reasonable and workable approach if expenditures of all governmental units were independent of national income. If, however, it is assumed, as is done above, that expenditures of state

and local governments are determined by tax revenues which are dependent on national income, then the traditional multiplier formula becomes unworkable. We cannot tell what figures to use for total governmental expenditures in the multiplier formula since a major component — state and local expenditures — is to be determined by the multiplier process itself. To make an estimate of total governmental expenditures for use in the usual formula, we would have to guess at the ultimate changes in national income and at the relationship between state and local expenditures and the national income. If we could guess the ultimate change in national income we would not need the multiplier formula at all. The proposed multiplier equation (31) confines itself to the assumedly predetermined element in total governmental expenditures, i.e. autonomous federal expenditures:

$$(31) \qquad Y = \frac{\dot{E}_a}{1 - \alpha - \lambda_a - \lambda_b - \lambda_c}$$

This formula does require estimates of the relationship between state and local expenditures and the national income, i.e. $\lambda_b$ and $\lambda_c$. It does not require, however, any guess at the ultimate level of national income which is implicitly involved in the use of (28).

To appreciate more fully the distinction between the two types of formulae, let us suppose that the magnitude of a contemplated change in federal expenditures, $E_a$, is known. No independent change in state and local expenditures, $E_b$ and $E_c$, is planned, but some change in state and local expenditures will result from the increased revenues associated with the higher national income. The formula, $\dfrac{X_a}{1 - \alpha}$, is inadequate because it leaves out the increases in income resulting from the state and local expenditures. We must somehow take account of induced changes in state and local expenditures. The easiest device would seem to be to add in the state and local expenditures, thus

$$(28) \qquad Y = \frac{X_a + X_b + X_c}{1 - \alpha} = \frac{X}{1 - \alpha}$$

This cannot be done because we do not know in advance what the state and local expenditures will be since they are dependent partly on the result to be obtained, the change in national income, $Y$. Instead of following this clumsy and inaccurate procedure, we confine ourselves to the independent change in federal expenditures, $E_a$, but alter the formula with respect to the use of the marginal propensity

to consume, $\alpha$. We add to $\alpha$ the fraction of increase in national income which the state and local governments spend, $\lambda_b$ and $\lambda_c$. The sum of these three, $\alpha + \lambda_b + \lambda_c$, is used in place of the marginal propensity to consume, $\alpha$, alone. Allowance is also made for any income-induced change in federal expenditures.

In the case of A, equation (20) implies virtually unlimited fiscal independence. Hence it may be assumed that the federal government might adopt a policy which automatically relates expenditures to fluctuations in the national income regardless of tax revenues. Thus the value of $\lambda_a$ may be derived from an examination of federal fiscal policy. For instance, the federal government may be following a policy of automatically cutting its expenditures by an amount equal to 10 per cent of any increase in national income. The value of $\lambda_a$ would then be $-.10$. In the case of state and local governments, however, it is reasonable to assume that expenditures are closely tied to tax revenues. In order to estimate $\lambda_b$ and $\lambda_c$, it is necessary, therefore, to examine first of all the budget policies of these governmental units. If, by and large, they follow a policy of balancing their budgets, as in equations (21) and (22), we may look to tax revenues for aid in estimating $\lambda_b$ and $\lambda_c$. Assuming that fluctuations in such revenues depend on national income, as in (15) and (16), we find that we may estimate $\gamma_b$ and $\gamma_c$ in order to obtain $\lambda_b$ and $\lambda_c$.

Modifications either in the assumptions concerning budgetary policy or in the behavior of state and local tax revenues in relation to income would upset the equalities $\gamma_b = \lambda_b$ and $\gamma_c = \lambda_c$, and would necessitate either seeking a direct estimate of $\lambda_b$ and $\lambda_c$ or determining some other relationship between the $\lambda$'s and the $\gamma$'s.

It is clear from formulae (31), (33), and (34) that if the $\lambda$'s are positive they reinforce the operation of the ordinary multiplier, i.e. increase the size of the multiplier. The change in federal expenditures and deficit (or surplus) required to achieve a given change in income is reduced. If the $\lambda$'s have different signs they work at cross-purposes, and the outcome will depend on the magnitude of the individual $\lambda$'s.

APPLICATIONS

It is interesting to apply the above formulae to a consideration of the error that is involved in failing to take account of state and local repercussions. If we neglect state and local repercussions,

$$(30) \qquad Y = \frac{E_a}{1 - \alpha - \lambda_a}$$

Therefore,

(35) $$E_a = Y\,(1 - \alpha - \lambda_a)$$

This gives us the change in federal expenditures required to achieve any specified change in income when state and local repercussions are ignored.  Since

(9) $$D_a = X_a - R_a$$
$$= E_a + N_a - R_a$$

(36) $$D_a = E_a + \lambda_a Y - \gamma_a Y$$

Substituting the value of $E_a$ given in (35),

(37) $$D_a = Y\,(1 - \alpha - \gamma_a)$$

This gives us the federal deficit required to achieve any specified increase in income.[1]  State and local repercussions are not taken into account.

Now if state and local repercussions are considered,

(33) $$Y = \frac{E_a}{1 - \alpha - \lambda_a - \lambda_b - \lambda_c}$$

Therefore,

(38) $$E_a = Y\,(1 - \alpha - \lambda_a - \lambda_b - \lambda_c)$$

This gives us the federal expenditures required to achieve any specified increase in income when state and local repercussions are considered.  From (36) and (38),

(39) $$D_a = Y\,(1 - \alpha - \gamma_a - \lambda_b - \lambda_c)$$

This gives us the federal deficit required to achieve any specified increase in income when state and local repercussions are considered. Comparing (35) with (38), we see that failure to take account of the state and local repercussions will exaggerate the needed federal expenditures to attain a given national income by

$$Y\,(1 - \alpha - \lambda_a) - Y\,(1 - \alpha - \lambda_a - \lambda_b - \lambda_c)$$
$$= Y\,(\lambda_b + \lambda_c)$$

i.e. by the amount of the change in state and local expenditures which results from the change in national income.

Comparing (37) with (39), we see that the same error is made if the

---

1. This type of formula is developed more fully in the section on "Expansive Effects of a Balanced Budget" in the writer's article on "The Impact of Fiscal Policy on National Income," *Canadian Journal of Economics and Political Science*, August 1942.  Cf. Paul A. Samuelson, "Simple Mathematics of Income Determination," in *Income, Employment and Public Policy: Essays in Honor of Alvin H. Hansen* (New York, 1948).

change in federal deficit is estimated without regard to state and local repercussions.

<div align="center">EXAMPLES</div>

It is assumed that $\alpha = .5$, $\gamma_a = .1$, $\lambda_a = -.2$

$\gamma_b = \lambda_b = .1$, $\gamma_c = \lambda_c = .2$

*Example I. State-local reinforcement of federal counter-inflationary policy.* Suppose that a reduction of \$10 billion is desired in money national income, i.e. $Y = -10$.

In the absence of state-local repercussions

$$
\begin{align}
(35) \qquad E_a &= Y (1 - \alpha - \lambda_a) \\
&= -10 (1 - .5 + .2) \\
&= -7
\end{align}
$$

$$
\begin{align}
(37) \qquad D_a &= Y (1 - \alpha - \gamma_a) \\
&= -10 (1 - .5 - .1) \\
&= -4
\end{align}
$$

Introducing state-local repercussions

$$
\begin{align}
(38) \qquad E_a &= Y (1 - \alpha - \lambda_a - \lambda_b - \lambda_c) \\
&= -10 (1 - .5 + .2 - .1 - .2) \\
&= -4
\end{align}
$$

$$
\begin{align}
(39) \qquad D_a &= Y (1 - \alpha - \gamma_a - \lambda_b - \lambda_c) \\
&= -10 (1 - .5 - .1 - .1 - .2) \\
&= -1
\end{align}
$$

The state-local reinforcement of federal policy makes it possible to achieve a reduction of \$10 billion in national income through a reduction of \$4 billion instead of a reduction of \$7 billion in federal expenditures. A reduction of \$7 billion in federal expenditures would bring about an excessive decline in national income. The reduction in deficit is only \$1 billion instead of \$4 billion.

*Example II. State-local reinforcement of federal counter-deflationary policy.* Suppose that an increase of \$10 billion is desired in money national income, i.e. $Y = 10$.

The same analysis holds as in Example I. In the absence of state-local repercussions,

$$
E_a = 7
$$
$$
D_a = 4
$$

Introducing state-local repercussions,

$$
E_a = 4
$$
$$
D_a = 1
$$

The state-local reinforcement of federal policy makes it possible to achieve an increase of \$10 billion in national income through an

increase of $4 billion instead of an increase of $7 billion in federal expenditures. Likewise, the deficit is increased by only $1 billion instead of being increased by $4 billion.

*Example III. State-local offsetting of federal counter-inflationary policy.* Suppose that autonomous public and private expenditures (not induced by income changes) are estimated to rise $10 billion. The federal government is following a drastic policy of automatically reducing its spending as national income rises, e.g. $\lambda_a$ has the high negative value assumed above, i.e. $- .2$.

In the absence of counter-inflationary federal policy and state-local repercussions, the money income would rise:

$$Y = 10 \cdot \frac{1}{1 - \alpha}$$

$$= 10 \cdot \frac{1}{1 - .5}$$

$$= 20$$

The federal policy would succeed in keeping this down to

$$Y = 10 \cdot \frac{1}{1 - \alpha - \lambda_a}$$

$$= 10 \cdot \frac{1}{1 - .5 + .2}$$

$$= 14.3$$

The induced increase in state-local expenditures, however, more than offsets this and permits money income to rise:

$$Y = 10 \cdot \frac{1}{1 - \alpha - \lambda_a - \lambda_b - \lambda_c}$$

$$= 10 \cdot \frac{1}{1 - .5 + .2 - .1 - .2}$$

$$= 25$$

Thus an even more drastic federal policy seems indicated, in view of the state-local repercussions, if the autonomous increase in expenditures is to be counteracted effectively. The state and local governments have reinforced the effects of the autonomous increase in expenditures.

*Example IV. State-local offsetting of federal counter-deflationary policy.* Suppose that autonomous public and private expenditures (not induced by income changes) are estimated to fall $10 billion and the federal government is following a drastic policy of automatically increasing its spending as national income falls, e.g. $\lambda_a$ has the high

negative value previously assumed, i.e. − .2. The algebraic analysis of Example *III* holds.

In the absence of counter-deflationary federal policy and state-local repercussions, the money income would fall $20 billion. The federal policy would make the reduction only $14.3 billion but the state-local governments are induced to cut their expenditures, too, with the result that income falls $25 billion. Again a more drastic federal policy is needed since the state and local governments have reinforced the effects of the autonomous decrease in expenditures.

One way of stating the results of these examples is to say that income-induced changes in state and local expenditures reinforce *autonomous* changes in federal expenditures (and all other autonomous changes in expenditures) but offset *income-induced* changes in federal expenditures where the latter are inversely related to income.

The above analysis has been confined to *income-induced* changes in state and local expenditures. *Autonomous* changes in state and local expenditures may, of course, reinforce or offset the effects of either autonomous or induced federal expenditures. The relatively limited fiscal powers of state and local governments suggest that changes in their expenditures are closely related to changes in national income and are, therefore, mainly income-induced rather than autonomous.

### CONCLUSIONS

Thus the federal and state-local governments may have mutually reinforcing or mutually offsetting effects on national income. An autonomous increase in federal expenditures may be reinforced by an income-induced increase in expenditures by state and local governments. An autonomous reduction in federal expenditures may be reinforced by an income-induced reduction in state and local expenditures. On the other hand, federal expenditures which are not autonomous (i.e. not independent of income changes) but which rise as national income falls may be offset by income-induced reductions in state and local expenditures. Federal expenditures which fall as national income rises may be offset by income-induced increases in state and local expenditures. These diverse relationships emphasize the importance of an adequate formulation of the multiplier relationship. Failure to recognize either of these opposing effects or failure to distinguish one type of effect from the other may lead to serious errors in income estimation and to either excessive or inadequate federal expenditures and deficit (or surplus) for any desired change in national income.

HAROLD M. SOMERS.

UNIVERSITY OF BUFFALO

## THE ECONOMIC REPORTS OF THE PRESIDENT
## AND THE PROBLEM OF INFLATION

The first Economic Report of the President, called for by the Employment Act of 1946, appeared in January 1947, and the fifth and most recent Report appeared in January 1949. These Reports have received unusually widespread public attention. The reprinting in a popular priced trade edition of the first three, and more recently of the five Reports[1] has made the full texts readily available and has contributed to a careful study of them in college class-rooms and by economists, bankers, business men and just plain citizens. It is true, as the Council of Economic Advisers points out in its Introduction to the reprint of the first three Reports, that "the ultimate judgments entering into the Economic Report to the Congress rest with the President," but the general public as well as economists believe that most of the technical analysis represents the professional opinion of the Council. It is therefore especially incumbent on economists to subject the Reports to a more searching criticism with respect to their economic analysis than would be warranted in the case of the ordinary public document originating in Congress or in the Executive Branch.

The Reports on two major points have taken a position with which most American economists would agree: the need to curb inflation, and (in the first three Reports) the desirability of not reducing taxes until a substantial surplus is assured and the inflationary situation is under control.[2] After tax reduction was enacted over the President's veto early in 1948, the two subsequent Reports continued to stress the importance of a budgetary surplus and to urge a selective tax increase.[3]

The fact that on the question of tax reduction professional economic opinion was opposed to the consensus in Congress and in

1. *The Economic Reports of the President.* Introduction by the Council of Economic Advisers. Harcourt, Brace and Co., 1949. All citations are to the 1949 reprint. In this reprint a new pagination is used after the 1949 Report, and hence citations to the 1949 Report are marked I, and citations to earlier Reports are marked II.

The individual reports were also published as House Documents and as separate Government Documents.

2. January, 1947, II, p. 34; July, 1947, II, pp. 61, 67, 85, 99; January, 1948, II, pp. 122, 163.

3. July, 1947, II, p. 257; January, 1949, I, p. xxiv.

business and financial circles, quite understandably would make many economists feel that the earlier Reports had analyzed correctly not only the seriousness of the inflationary danger in the period covered but also the nature of the forces back of price movements. But more critical reading of the Reports fails to support the favorable first impression. Not only do they, in particular the earlier ones, have many statements about the price situation that in effect run contrary to their expressed fears about inflation and the inflationary dangers in tax reduction, but some of the economic analysis of the price situation is open to serious question.

A comparison of the Report of January 1947 with that of a year later reveals much less concern over inflation in 1947 than in 1948. In fact, the first Report when considered as a whole, despite some general statements about the dangers of inflation and the evils of high prices, gives the impression that production and employment were in a precarious situation, maintained only by the temporary forces of domestic and international investment. In the face of an increase in 1946 of nearly twenty per cent in the cost of living, and of over thirty per cent in food prices and in wholesale prices, the whole trend of the discussion is on the necessity of maintaining a high level of monetary expenditure. This is illustrated by the analysis in Part V, "Favorable and Unfavorable Factors in 1947,"[4] where the forces that might hold up prices or send them still higher are classed as "favorable," and those that might keep prices from going higher or might lower them are classed as "unfavorable." "To maintain maximum production and employment in 1947 it is desirable that business investment be at an annual rate of at least equal to the annual rate prevailing in the last quarter of 1946."[5] Exports in 1946 "played an important role in the maintenance of domestic production, employment, and purchasing power and may be expected to do so this year."[6] The availability of "abundant aggregate funds including ample bank credit" is a "clear element of strength" in business demand.[7]

That economists at the end of 1946 should have misjudged the nature of the price developments in the next two years is understandable. Few if any economists have a perfect record of prediction. But an examination of the analysis in both the 1947 and 1948 Reports suggests that the misjudging of the situation was not simply the misplaced emphasis or the bad timing that can upset the predictions of

4. II, pp. 21–28.
5. II, p. 24.
6. II, p. 44.
7. II, p. 27.

even the ablest of economic analysts, but was inherent in the fallacious conception in the Reports of the nature of inflationary and deflationary forces in the modern economy. In the January 1948 Report the inflationary danger is given more prominence,[8] and the Midyear Report of 1948 states that "we are in the very midst of gathering inflationary forces." "The favorable factors" of the 1947 Report — consumer demand, business demand, exports, and an easy credit situation — have eighteen months later become part of the "interacting processes making for continued inflation."[9] However, insofar as the 1948 Reports represent a more realistic appraisal of the price situation, the realism results largely from being faced by the facts of a continuing upward movement of prices rather than from a re-examination of the analysis of the previous year.[1]

In brief, the Reports of 1947 and 1948 suffer from four analytical errors:

1. The assignment of a very minor role to, and in places the almost complete ignoring of, the increase of the money supply as a factor in the inflationary picture, and the repeated implications that rising prices are due in large part to the unreasonable action of business men and of labor leaders.

2. An assumption that the volume of production and employment has an almost 100 per cent correlation with money payments, even in a period of practically full employment.

3. The failure to recognize that, in a period of practically full employment, capital formation and foreign investment must of necessity reduce the product available for consumption.

4. An unwarranted alarm over the recent decline in the purchasing power of per capita disposable income, as a result of the failure to recognize that a basic reason for the decline in real disposable income has been the impact of domestic and foreign investment demand, and of an increasing propensity to consume, on the price structure.

*Neglect of Monetary Causes of Inflation*

The shortcomings of the Reports on this point are ones of emphasis rather than of complete neglect, for even in the early ones

8. II, pp. 157–160.
9. II, p. 255.
1. In the 1949 Report, following the small drop in a number of price indices in the closing months of 1948, there is a shift in emphasis to the monetary basis of inflation, with the result that the economic analysis in this Report is free from some of the defects in the earlier ones.

there are references to monetary and credit conditions.[2]   Not until
the January 1948 Report, however, is the monetary situation dis-
cussed as an important factor related to inflation,[3] but neither in
that discussion nor anywhere else previous to the January 1949
Report is there any reference to the great expansion in the monetary
supply since 1939.   The 1949 Report says that money and credit
"increased enormously during the war,"[4] but up to that time nothing
in the text of the Reports would give the reader any inkling of the
fact that currency in circulation in 1947 was over four and a quarter
times what it was eight years before, or that demand deposits adjusted
and currency combines were three and a half times what they were
eight years before.

The reader of the Reports gets the impression that the heart
of the problem of inflation is in the attitude of business men and
labor leaders.   These men are repeatedly exhorted to be reasonable.
The first Report lectures business and labor on the virtues of moder-
ation.[5]   The second Report points out that the President, after the
abandonment of price control, had "urged businessmen to resist
inflationary pressures or temptation and to make voluntary price
reductions wherever possible;" and gives similar advice to labor
about wage increases.[6]   The January 1948 Report reiterates the
theme, urging businessmen to "hold the line against price increases
and reduce prices wherever they can," and preaching that "labor
should be moderate in its wage demands, mindful of recent experience
which demonstrates the impossibility of registering real gains in an
inflationary spiral."[7]   In the 1948 Midyear Report, with what
seems like unintended irony in view of the alarm shown in that Report
over inflation, we are told that "many leaders in both industry and
labor can be applauded for the conscious restraint they have exhibited
in their pricing policies and wage demands."[8]

*Relations of Monetary Payments to Physical Production
   and Employment*

On this point passages can be found to support almost any view,
but the dominant theme, particularly in the earlier Reports, is that

2. January, 1947, II, p. 10.
3. II, pp. 164–66.   In the January 1947 Report, "ample bank credit" had
been mentioned under the heading of "favorable factors in business demand,"
II, p. 27.
4. I, p. xix.
5. II, pp. 31, 32.
6. II, pp. 89, 96.
7. II, p. 168.
8. II, p. 255.

physical production and employment move up and down in almost exact relation to money payments. In the first 1947 Report appears the categorical statement: "The volume of employment and production in any given period depends upon the volume of expenditures."[9] Foreign sales in 1946 "played an important role in the maintenance of domestic production, employment, and purchasing power and may be expected to do so this year."[1]

The Midyear Report for 1947 is on the defensive in the matter of the export surplus, but the emphasis is on the quantity of goods available for the consumer rather than upon the price consequences, although there is a reference to foreign buyers' contribution to the "upward pressure on prices."[2] Instead of praising the export surplus as a stimulant to production, as was done six months earlier, the point is now made that "it is easy to exaggerate the degree of deprivation" for home use.[3] In itself this is a correct statement, but it contributes to the misleading picture that foreign demand in 1947 had its impact largely upon production and not upon prices. In the 1948 Report there is a further shift in emphasis, and the Report states that exports in the first half of 1947 "had an important bearing upon the total levels of employment and production and upon the level of prices."[4] It is certainly open to question, however, whether exports had an appreciable effect upon total production and employment, particularly in view of the fact that in the case of many manufactured and agricultural products, exports simply deprived domestic consumers of part of an output which was at a maximum.

*Relation Between Consumer Income, Capital Formation,*
*and Foreign Aid*

The first Report stresses the point that appears in subsequent ones, namely, that per capita real "disposable income" (personal income after taxes), after rising in 1944 to more than 50 per cent above the 1935–39 average, had fallen slightly in 1945 and 1946, and that only by a substantial reduction in savings (from 28.3 per cent of disposable income in 1944 to 9.5 per cent of disposable income in the last quarter of 1946) and by an increase in installment credit

9. II, p. 14.
1. II, p. 44. As "purchasing power" is defined in the Reports in real terms (II. p. 10), this statement would appear to mean that in this period of virtually full employment the product available to the American economy was greater as a result of our large exports.
2. II, pp. 70, 111.
3. II, p. iii.
4. II, p. 144.

had it been possible for consumers to purchase the same quantity of goods and services as before.[5]   The January 1948 Report observes that with the further decline in the real income of consumers in 1947, consumption "was maintained by liquidation of past savings, a reduction of the saving out of current incomes, and the extensive use of consumer credit,"[6] and the 1949 Report points out that in 1948 there was no appreciable gain in consumer per capita real income despite a rise in total national output.[7]

When an economy reaches approximately full employment, particularly if there is a decline in overtime work and a stoppage of production through strikes, it is not surprising that less goods will be available for consumers, if business is to increase capital plant and if we are to have a large export surplus, whether that export surplus be financed by credits, gold shipments to the United States, or a drawing down of dollar balances.   The economic miracle of the years 1940 to 1944, of rising real per capita disposable income in the face of a diversion of a large proportion of production to ends other than consumption, requires for its accomplishment the stage property of a large volume of unemployment at the start of the performance. The Reports seem to miss this point and to consider the development of the years 1940 to 1944 as a normal feature of the American economy even in a state of full employment.   They have an almost mystical belief that any decline in real per capita disposable income is a threat to economic stability, regardless of the cause of the decline.

Insofar as consumers, by reducing their rate of savings, by borrowing, or by drawing on past savings, refuse to adjust their consumption to their reduced real disposable income, the effect must be either that goods are diverted from investment to consumption, or that the allocation to consumption is unchanged but on the basis of higher prices, or that there is some combination of these effects dependent upon the various price, income, and supply elasticities. As a result of the combination of a belief in the normality of an increase in per capita real disposable income, and the assumption — perhaps not recognized explicitly but nevertheless basic to much of the analysis — that the problem of the inflationary years of 1946 and 1947 are much the same as the problems of the 1930s, the Reports are confused as to the significance of this rising propensity to consume, and as to its desirability in the period covered.   This confusion is illustrated by the discussion in the January 1948 Report

5. II, pp. 21–25.
6. II, p. 131.
7. I, p. xix.

where the "survey of consumer income and expenditures during 1947 points to two causes for concern:" the contribution to "inflationary pressure," and a reduction in "the real purchasing power of consumers to the point where it will not be adequate to take the consumers' share of the national output at present prices when certain abnormal demands are reduced and when the rate of consumer saving cannot be further reduced."[8]

*The Probable Consequences of a Decline in Demand from Domestic Capital Formation and Export Surplus*

The authors of the Reports have got themselves into a vicious circle of their own creation on this whole problem of the relation of consumer purchasing power, investment demand, and export surplus, to prices and employment. In individual passages, particularly in the 1948 Reports, the influence of these demands upon prices is recognized,[9] but the significance of this inflationary impact is never integrated into the analysis. The inflationary effects of increased expenditures of a temporary type are regarded as non-reversible. The thesis of the Reports seems to be that although an increase in money payments from domestic capital formation, export surplus, and a high propensity to consume may raise prices and distort previously existing price relations, a decline in capital formation, a smaller export surplus, and a lower propensity to consume will not bring a downward movement of any individual prices. Given these assumptions, the conclusion of course follows that a slackening of demands from these three sources must result in depression and unemployment. The real issue is whether the Reports' approach to the problem is a realistic one, or one that will contribute to a solution of the problem of readjustment with the slackening of these demands.

Repeatedly the point is made that if these special demands fall off, consumer income cannot absorb output at existing prices,[1] but with the exception of one statement in the 1949 Report,[2] I find nothing that suggests that the end of these special demands will bring any pressure toward a readjustment of price relations, much less that forces of the market will actually bring about such a read-

8. II, pp. 136–137.
9. II, pp. 157–158. "The inflationary impact of spending was strong in all fields — capital expansion, exports, and consumption." The same idea appears in the Midyear Report for 1948, II, p. 255.
1. January, 1947, II, pp. 20, 30–31; July, 1947, II, pp. 90, 92, 101, 108; January, 1948, II, pp. 119, 125, 137, 149, 195; January, 1949, I, pp. xvii, xxi.
2. I, p. xvii.

justment. The position of the Reports on such an adjustment seems to be that business men should become more reasonable or that the government should take more vigorous action against inflation, and not that there are forces in the market which would make such an adjustment likely.[3]

I would not minimize the reality of the problem of adjustment of production and of prices with a decline in the demand for individual products. A sound economic policy must not only obviate drastic decreases in total money payments, but it should also recognize that in some areas the prices of individual products will decrease in response to a decrease in money payments, and that the extent of these price decreases, and the area over which they operate, is to some degree dependent upon governmental policy. The failure of the Reports to recognize this twofold nature of the problem of adjustment to changing monetary demand, and their consequent assumption that unemployment must be the result of any decrease in money payments — except as prices are reduced from causes that apparently are not connected with the decrease in money payments — has resulted from their circular reasoning on the relationship between the various parts of the national product.

The Reports from the beginning have performed a public service in stressing inflationary dangers and in pleading for the unpopular policy of maintaining high taxes. They contain much excellent economic analysis, but the thinking of the authors is permeated with the idea that real disposable incomes of consumers must rise, regardless of other developments, and much of the discussion is shackled to the fears of deflation and mass unemployment. The neglect of the monetary basis of inflation, the attempt to make business men and labor leaders personally responsible for inflation, and the repeated assumption that the impact of demand changes is almost exclusively upon production and employment rather than upon prices, results in an unbalanced picture of the forces of inflation and of deflation.

The Reports raise for economists the larger question as to the use and the limitations of logical analysis as a basis for prediction and as a basis for the formulation of public policy. The dominant thought in them is Keynesian, in the broad sense of the body of doctrine that was inspired by the *General Theory*. Underlying most of the analysis is the basic assumption that changes in aggregate demand have employment effects rather than price effects, and that a steady increase in consumer purchasing power is essential to

3. January, 1949, I, p. xxi; January, 1947, II, pp. 20, 21, 30, 31; July, 1947, II, p. 88.

economic stability with full employment. A great deal of the controversy of the last thirteen years over consumption, savings, investment, aggregate demand, and employment, that is associated with the name of Keynes, has been a battle of logicians. Interpreted with perspective, the fruits of this controversy have contributed greatly to the understanding of the monetary economy. Yet those on both sides of the argument have all too frequently taken the view that logic could settle public policy. Many ardent Keynesians, acutely conscious of the fact that they have discarded the conclusions of Ricardo and his followers, apparently do not realize that their conceptual approach is almost identical with that of Ricardo and of other great figures of the classical tradition: there are economic laws which when mastered tell us how the economic system behaves at all times.

In the later 1930's the Keynesian approach seemed to many economists — and certainly to a great majority of younger economists — a more satisfying explanation as to how the modern economy had behaved in the last few years than was anything found in classical economics. What was all too frequently forgotten, or greatly underestimated, was that the validity of either approach depends in large part on such shifting factors as the current psychology of consumers, the nature of the pricing practices affecting individual products, and the existing investment situation. In view of this situation, economic theory, if it is to be useful for prediction, must be considered as a hypothesis whose applicability in any particular situation depends upon the changing human reactions to particular monetary impulses. A failure to recognize the limitations of pure analysis, and the assumption that conditions in 1946 were much like those in the early 1930's have been the basic shortcomings in the Reports' discussion of inflation and employment.

FRANK WHITSON FETTER.

NORTHWESTERN UNIVERSITY

# RECENT PUBLICATIONS

ALEXANDER, R. S., SURFACE, F. M. AND ALDERSON, W. Marketing. Boston, Mass.: Ginn and Company. 1949. pp. 840. $5.75. Revised edition.

THE ASSOCIATION OF CERTIFIED AND ASSOCIATE ACCOUNTANTS. Working Party Reports. London, England: Library and Publications Committee. 1948. pp. 72.

BAKER, HELEN. Company-Wide Understanding of Industrial Relations Policies: A Study in Communications. Princeton, New Jersey: Industrial Relations Section, Princeton University. 1948. pp. 79. $2.00.

BANK FOR INTERNATIONAL SETTLEMENTS. Reglementation du Commerce de l'Or et des Devises en France, Supplement 5. Basle, Switzerland: The Bank. 1948. pp. 50.

BERRIDGE, WILLIAM A. AND WOLFE, CEDRIC. Guaranteed Employment and Wage Plans. Washington, D. C.: American Enterprise Association, Inc. 1948. pp. 87. 50 cents.

BINNS, K. J. Federal Financial Relations in Canada and Australia. Hobart, Tasmania, Australia: Government Printer. 1948. pp. 75.

BOGART, ERNEST L. Peacham, the Story of a Vermont Hill Town. Montpelier, Vermont: Vermont Historical Society. 1948. pp. 494. $6.00.

BRADFORD, FREDERICK A. Money and Banking. New York: Longmans, Green and Company, Inc. 1949. pp. 860. $4.75. [Sixth Edition.]

BRISCOE, VERA; MARTIN, JAMES W.; REEVES, J. E. Safeguarding Kentucky's Natural Resources. Lexington, Kentucky: University of Kentucky, Bureau of Business Research and the Bureau of Government Research. 1948. pp. 224.

BURGESS, GEORGE H. AND KENNEDY, MILES C. Centennial History of the Pennsylvania Railroad Company. Philadelphia, Penn.: The Pennsylvania Railroad Company. 1949. pp. 835. $3.50.

CANADA, GOVERNMENT OF. Investment and Inflation, with Special Reference to the Immediate Post-War Period, Canada, 1945–1948. Ottawa, Canada: The Canadian Government. 1949. pp. 290.

CARTER, C. F., REDDAWAY, W. B. AND STONE, RICHARD. The Measurement of Production Movements. Cambridge, England: Cambridge University Press. 1948. pp. 135. 12s. 6d. net. [Department of Applied Economics Monograph 1, Cambridge University.]

CASSADY, RALPH AND JONES, Wylie I. The Changing Competitive Structure in the Wholesale Grocery Trade. Berkeley, California: University of California Press. 1949. pp. 73.

COCHRAN, THOMAS C. The Pabst Brewing Company: The History of an American Business. New York: New York University Press. 1948. pp. 451. $5.00.

COHEN, SANFORD. State Labor Legislation, 1937–1947. Columbus, Ohio: Bureau of Business Research, Ohio State University. 1949. pp. 150. $2.50.

THE COMMITTEE ON PUBLIC DEBT POLICY. Our National Debt: Its History and Its Meaning Today. New York: Harcourt, Brace and Company. 1949. pp. 182. $2.50.

COPE, ALFRED HAINES. The Administration of Civil Service in Cities of the Third Class in Pennsylvania. Philadelphia, Penn.: The University of Pennsylvania. 1948. pp. 166.

THE COUNCIL OF ECONOMIC ADVISERS. The Economic Reports of the President. New York: Harcourt, Brace and Company. 1949. pp. 333. $2.75.

COUNCIL OF KING'S COLLEGE. John Harold Clapham, 1873–1946. Cambridge, England: The College. 1949. pp. 28. 3s. 6d.

COUNCIL OF KING'S COLLEGE. John Maynard Keynes, 1883–1946. Cambridge, England: The College. 1949. pp. 42. 3s. 6d.

DEPARTMENT OF INDUSTRIAL RELATIONS, LAVAL UNIVERSITY. Formes de Collaboration Patronal-Ouvrière. Quebec, Canada: The Faculty of Social Sciences. 1948. pp. 186. $1.50.

DORFMAN, JOSEPH. The Economic Mind in American Civilization, Volume III, 1865–1918. New York: The Viking Press, Inc. 1949. pp. 494. $6.00.

DUERR, WILLIAM A. The Economic Problems of Forestry in the Appalachian Region. Cambridge, Mass.: Harvard University Press. 1949. pp. 308. $5.00. [Harvard Economic Studies No. 84.]

DUNLOP, JOHN T. Collective Bargaining, Principles and Cases. Chicago, Illinois: Richard D. Irwin, Inc. 1949. pp. 433. $4.00.

EDWARDS, CORWIN D. Maintaining Competition. New York: McGraw-Hill Book Company. 1949. pp. 337. $3.75.

EITEMAN, WILFORD J. Price Determination: Business Practice versus Economic Theory. Ann Arbor, Michigan: School of Business Administration, University of Michigan. 1949. pp. 39. $1.00.

EVANS, GEORGE HEBERTON, JR. Business Incorporations in the United States. New York: National Bureau of Economic Research, Inc. 1948. pp. 184. $6.00.

GADGIL, D. R. Some Observations on the Draft Constitution. Poona, India: Gokhale Institute of Politics and Economics. 1948. pp. 112. Rs. 3 or 4s. 6d. [Publications No. 19.]

GALBRAITH, JOHN KENNETH. Beyond the Marshall Plan. Washington, D. C.: National Planning Association. 1949. pp. 28. 50 cents.

GALENSON, WALTER. Labor in Norway. Cambridge, Mass.: Harvard University Press. 1949. pp. 375. $5.00.

GIBSON, ROLAND. Cotton Textile Wages in the United States and Great Britain: A Comparison of Trends, 1860–1945. New York: King's Crown Press. 1948. pp. 137.

GIDE, CHARLES, AND RIST, CHARLES. A History of Economic Doctrines from the Time of The Physiocrats to the Present Day. Boston: D. C. Heath and Company. 1948. pp. 800. $7.00. 7th Edition.

GRAHAM, HERMAN D. The Economics of Strip Coal Mining. Urbana, Illinois: University of Illinois Press. 1948. pp. 77.

GRAHAM, FRANK D. The Cause and Cure of "Dollar Shortage." Princeton, New Jersey: Princeton University, International Finance Section. 1949. pp. 15. [Essays in International Finance, No. 10.]

GRUBER, KARL. Die Zusammenhänge zwischen Grösse, Kosten und Rentabilität Industrieller Betriebe. Vienna: Springer. 1948. pp. 24.

HANSEN, A. H. Monetary Theory and Fiscal Policy. New York: McGraw-Hill Book Company. 1949. pp. 236. $3.00. [Economics Handbook Series.]

HAWAIIAN ECONOMIC FOUNDATION. A Study of Ownership of Corporations in Hawaii. Honolulu, Hawaii: The Foundation. 1948. pp. 35.

HEINIG, KURT. Das Budget. Stockholm, Sweden: Paul Siebeck. 1949. pp. 320. DM 8.-

HICKS, J. R. The Problem of Budgetary Reform. Oxford, England: Clarendon Press. 1948. pp. 95. $1.50.

HOLBORN, LOUISE W. War and Peace Aims of the United Nations, Volume II, 1943–1945. Boston, Mass.: World Peace Foundation. 1949. pp. 1278. $6.00.

HOPKINS, EDWIN. The High Crime of Inflation: How to Restore the Purchasing Power of the Dollar. New York: The Author. 1948. pp. 217. $4.00.

HOPKINS, JOHN A. AND HEADY, EARL O. Farm Records. Ames, Iowa: Iowa State College Press. 1949. pp. 303. $3.75.

HOPKINS, WILLIAM STEPHEN. Labor in the American Economy. New York: McGraw-Hill Book Company. 1949. pp. 368. $3.50.

HULTGREN, THOR. American Transportation in Prosperity and Depression. New York: National Bureau of Economic Research, Inc. 1949. pp. 397. $5.00.

HUTCHINS, OPPENHEIMER, CARR, ET AL. Foundations for World Order. Denver, Colorado: University of Denver Press. 1949. pp. 174. $3.00.

INSTITUT NATIONAL DE LA STATISTIQUE ET DES ETUDES ECONOMIQUES. L'Economie du Plein Emploi. Paris, France: Presses Universitaires de France. 1948. pp. 234.

INSTITUT NATIONAL DE LA STATISTIQUE ET DES ETUDES ECONOMIQUES. La Palestine. Paris, France: Presses Universitaires de France. 1948. pp. 183.

INSTITUT NATIONAL DE LA STATISTIQUE ET DES ETUDES ECONOMIQUES. Le Marché Mondial du Coton. Paris, France: Presses Universitaires de France. 1948. pp. 270. 600 francs.

INTER-ALLIED REPARATION AGENCY. Report of the Secretary General for the Year 1948. Brussels, Belgium: The Agency. 1949. pp. 98.

INTERNATIONAL LABOUR OFFICE. Housing and Employment. Geneva, Switzerland: The Office. 1948. pp. 147. 75 cents.

INTERNATIONAL LABOUR OFFICE. Second Session of the Permanent Migration Committee. Geneva, Switzerland: The Office. 1948. pp. 172. $1.00.

IOWA STATE TAX COMMISSION. Annual Report for the Period from July 1, 1947, to June 30, 1948. Des Moines, Iowa: The Commission. 1949. pp. 352.

JACOBY, ERICH H. Agrarian Unrest in South East Asia. New York: Columbia University Press. 1949. pp. 287. $4.00.

KALDOR, NICHOLAS, AND SILVERMAN, RODNEY. A Statistical Analysis of Advertising Expenditure and the Revenue of the Press. Cambridge, England: Cambridge University Press. 1948. pp. 200. 15s. net. [National Institute of Economic and Social Research, Economic and Social Studies, VIII.]

KEIRSTEAD, B. S. The Theory of Economic Change. New York: The Macmillan Company. 1949. pp. 386. $5.00.

KENNEDY, THOMAS. Effective Labor Arbitration: The Impartial Chairmanship of the Full-Fashioned Hosiery Industry. Philadelphia, Penn.: University of Pennsylvania. 1948. pp. 286.

KROULD, HARRY J. Reference Notes on the Press in European Countries Participating in the European Recovery Program. Washington, D. C.: The Library of Congress. 1948. pp. 39. 25 cents.

LACOUR-GAYET, JACQUES, AND LACOUR-GAYET, ROBERT. De Platon a la Terreur. Paris, France: SPID. 1949. pp. 268. 300 fr.

LA VOLPE, GIULIO. Convenienza Economica Collettiva. Padova, Venice, Italy: Facolta di Economia e Commercio, Collana ca Foscari. 1948. pp. 281.

LAZARUS, HERMAN; AND GOLDBERG, JOSEPH P. Collective Bargaining. Washington, D. C.: The Public Affairs Institute. 1949. pp. 71. 50 cents.

LEWIS, EDWIN HENDERSON. Marketing Patterns of Philadelphia Wholesalers. Philadelphia, Penn.: University of Pennsylvania. 1948. pp. 90.

THE LIBRARY OF CONGRESS, EUROPEAN AFFAIRS DIVISION. The United States and Postwar Europe: A Bibliographical Examination of Thought Expressed in American Publications during 1948. Washington, D. C.: The Library. 1948. pp. 123.

LITTERER, OSCAR F. Where Does Small Business Obtain its Capital? Minneapolis, Minnesota: Federal Reserve Bank of Minneapolis. 1948. pp. 18.

LONIGAN, EDNA. Expanding Welfare in a Free Economy. Washington, D. C.: American Enterprise Association, Inc. 1949. pp. 44. [No. 431 in the series, "National Economic Problems."]

LÖWEGREN, GUNNAR. Swedish Iron and Steel: A Historical Survey. Stockholm, Sweden. Svenska Handelsbanken. 1948. pp. 123.

LUCK, DAVID J. Consumer Buying Habits in the Crab Orchard Area of Illinois. Urbana, Illinois: The University of Illinois. 1948. pp. 30. [Business Studies Number 8.]

LUCK, DAVID J. Operating Results of Illinois Retail Jewelry Stores in 1947. Urbana, Illinois: University of Illinois. 1948. pp. 31. Business Studies No. 7.

MACLAURIN, W. RUPERT. Invention and Innovation in the Radio Industry. New York: The Macmillan Company. 1949. pp. 304. $6.00.

MAHR, ALEXANDER. Volkswirtschaftslehre. Vienna, Austria: Springer. 1948. pp. 393.

MATHERLY, ALVA MARIAN. Kentucky State Purchasing. Lexington, Kentucky: The University of Kentucky. 1948. pp. 61. 50 cents.

MINISTÈRE DES AFFAIRES ETRANGÈRES. Documents Diplomatiques Francais (1871–1914) Tomes IX and X. Paris, France: Imprimerie Nationale. 1948. pp. 446 (IX), pp. 742 (X).

MONTANER, ANTONIO. Der Institutionalismus als Epoche Amerikanischer Geistesgeschichte. Tübingen, Germany: J. C. B. Mohr (Paul Siebeck). 1948. pp. 155.

MOSSÉ, ROBERT. Le système Monétaire de Bretton Woods et les Grands Problèmes de l'Après-Guerre. Paris, France: Recueil Sirey. 1948. pp. 152.

MOULTON, HAROLD G. Economic Systems. Washington, D. C. The Brookings Institution. 1948. pp. 55. 50 cents.

NATIONAL BUREAU OF ECONOMIC RESEARCH, INC. Studies in Income and Wealth, Volume 11. New York: The Bureau. 1949. pp. 450. $6.00.

NATIONAL INDUSTRIAL CONFERENCE BOARD. The Economic Almanac for 1949. New York: The Board. 1949. pp. 560.

NATIONAL PLANNING ASSOCIATION. 1949 Report to Members: Planning by Americans in Agriculture, Business, Labor, the Professions. Washington, D. C.: The Association. 1949. pp. 36.

THE NPA AGRICULTURE COMMITTEE ON NATIONAL POLICY. Must We Have Food Surpluses? Washington, D. C.: National Planning Association. 1949. pp. 46. 50 cents.

NELSON, LOWRY. Can Farmers Afford to Live Better. Washington, D. C.: National Planning Association. 1949. pp. 32. 50 cents. [Planning Pamphlets No. 65.]

NORTHRUP, F. S. C., Editor. Ideological Differences and World Order. New Haven, Connecticut: Yale University Press. 1949. pp. 486. $4.50.

NOVICK, DAVID; ANSHEN, MELVIN L.; AND TRUPPNER, WILLIAM C. Wartime Production Controls. New York: Columbia University Press. 1949. pp. 441. $6.00.

OEHLER, CHRISTIAN. Audits and Examinations: Standards and Procedures. New York: American Book Company. 1949. pp. 570. $5.25.

O'HARA, JAY L. Money and Banking. New York: Pitman Publishing Company. 1948. pp. 671. $4.75.

ORGANIZATION FOR EUROPEAN ECONOMIC COOPERATION. History and Structure of the Organization. Paris, France: The Organization. 1948. pp. 142.

ORGANIZATION FOR EUROPEAN ECONOMIC COOPERATION. Interim Report on the European Recovery Program. Paris, France: The Organization. 1948. pp. 195.

ORGANIZATION FOR EUROPEAN ECONOMIC COOPERATION. Report to the Economic Cooperation Administration on the 1949–1950 Programme, Volume I. Paris, France: The Organization. 1948. pp. 133.

OSBORN, RICHARDS C. The Renegotiation of War Profits. Urbana, Illinois: The University of Illinois. 1948. pp. 112. [Bureau of Economic and Business Research Bulletin Series Number 67.]

OUREN, TORE, AND SØMME, AXEL. Trends in Inter-War Trade and Shipping. Bergen, Norway: J. W. Eides. 1948. pp. 72.

OXFORD ECONOMIC PAPERS, WILSON, T. (Editor). Oxford, England: Oxford University Press. 1949. pp. 144. 12s. 6d. [New Series: Volume 1, Number 1.]

PARDU, BEULAH LEA. State Supervision of the Property Tax Assessments in Kentucky. Lexington, Kentucky: University of Kentucky. 1948. pp. 67. 50 cents.

PARES, BERNARD. Russia: Its Past and Present. New York: The New American Library of World Literature, Inc. 1949. pp. 224. 35 cents.

PETERSON, J. MARVIN, AND CAWTHORNE, D. R. Money and Banking. New York: The Macmillan Company. 1949. pp. 630. $5.00.

PETERSON, SHOREY. Economics. New York: Henry Holt & Company. 1949. pp. 918. $4.60.

PHELPS BROWN, E. H. The Framework of the Pricing System. Lawrence, Kansas: Student Union Book Store. 1949. pp. 221. $3.25. [First published by Chapman & Hall, Ltd., Oxford, England.]

PIGOU, A. C. Il Reddito Nazionale. Milan, Italy: L'Industria. 1949. pp. 103. 600. lire.

PIGOU, A. C. The Veil of Money. New York: The Macmillan Company. 1949. pp. 150. $2.50.

PIROU, GAËTAN. La Valeur et Les Prix. Paris, France: Recueil Sirey. 1945–46. pp. 533.

PIVER, J. C. Sun and Mankind. New York: The Author. 1949. pp. 58.

RATCLIFF, RICHARD U. Urban Land Economics. New York: McGraw-Hill Book Company, Inc. 1949. pp. 533. $5.50.

ROHLFING, CHARLES; CARTER, EDWARD; AND WEST, BRADFORD. Business and Government. Brooklyn, New York: The Foundation Press, Inc. 1949. pp. 808. $5.50.

ROSTAS, L. Comparative Productivity in British and American Industry. Cambridge, England: Cambridge University Press. 1948. pp. 263. 18s. net. [Occasional Papers XIII, of National Institute of Economic and Social Research.]

THE ROYAL CANADIAN INSTITUTE. The Royal Canadian Institute Centennial Volume. Toronto, Ontario: The Institute. 1948. pp. 232. Edited by: W. Stewart Wallace.

RUGGLES, RICHARD. An Introduction to National Income and Income Analysis. New York: McGraw-Hill Book Company. 1949. pp. 349. $3.75.

SANDAGE, CHARLES H. Qualitative Analysis of Radio Listening in Two Central Illinois Counties. Urbana, Illinois: Bureau of Economic and Business Research. 1949. pp. 57. [Bulletin No. 68.]

SARDA, JUAN. La Politica Monetaria y las Fluctuaciones de la Economia espanola en el siglo XIX. Madrid, Spain: Instituto Sancho de Moncada de Economia. 1948. pp. 261.

SCHULTZ, WILLIAM J. Social Security and the Economics of Saving. New York: National Industrial Conference Board. 1949. pp. 72. [Studies in Individual and Collective Security No. 5.]

SCHUMANN, C. G. W., AND SCHEURKOGEL, A. E. Industrial and Commercial Share Price Indices in South Africa. Union of South Africa: University of Stellenbosch. 1948. pp. 97.

SEERS, DUDLEY. Changes in the Cost-of-Living and the Distribution of Income since 1938. Oxford, England: Basil Blackwell. 1949. pp. 84. 6s.

SHISTER, JOSEPH. Economics of the Labor Market. Chicago, Illinois: J. B. Lippincott Company. 1949. pp. 590.

SIMONS, HENRY. Economic Policy for a Free Society. Chicago 37, Illinois: University of Chicago Press. 1948. pp. 353. $3.75.

SLICHTER, SUMNER H. The American Economy. New York: Alfred Knopf, Inc. 1949. pp. 214. $2.75.

SMITH, R. ELBERTON. Customs Valuation in the United States. Chicago, Illinois: The University of Chicago Press. 1948. pp. 380. $7.50.

THE SMITHSONIAN INSTITUTION. Annual Report of the Board of Regents, showing the Operations, Expenditures, and Condition of the Institution for the Year Ended June 30, 1947. Washington, D. C.: Government Printing Office, 1948. pp. 470. $2.00.

SONNE, HANS CHRISTIAN.   Enterprise Island.   New York: Business Bourse.
    1949.  pp. 128.  $3.50.
SOVANI, N. V., Editor.   Reports of the Commodity Prices Board.   Poona 4,
    India: Gokhale Institute of Politics and Economics.   1948.   pp. 239.
    15s. or 10 Rs.  [Publication No. 20.]
SOVANI, N. V.   The Social Survey of Kolhapur City: Part 1 — Population and
    Fertility.  Poona, India: Gokhale Institute of Politics and Economics.  1948.
    pp. 83.  Rs. 4 or 8s.  [Publication No. 18.]
SPIEGEL, HENRY WILLIAM.   The Brazilian Economy: Chronic Inflation and
    Sporadic Industrialization.   Philadelphia: The Blakiston Company.   1949.
    pp. 246.  $4.50.
SUTHERLAND, EDWIN H.   White Collar Crime.   New York: The Dryden Press.
    1949.  pp. 266.  $3.00.
SWEENEY, FRANK.   The European Economic Situation — 1948.   Washington,
    D. C.: American Enterprise Association, Inc.   1949.   pp. 40.   50 cents.
    [No. 429 in the Series "National Economic Problems."
SWEEZY, PAUL M.   Socialism.   New York: McGraw-Hill Book Company.   1949.
    pp. 276.  $3.50.  [Economics Handbook Series.]
THOMSON, CHARLES A. H.   Overseas Information Service of the U. S. Govern-
    ment.   Washington, D. C. The Brookings Institution.   1948.   pp. 397.
    $4.00.
TONIOLO, GIUSEPPE.   Capitalismo e Socialismo.   Citta del Vaticano, Rome,
    Italy: Comitato Opera Omnia di G. Toniolo.   1947.   pp. 527.
ULMER, MELVILLE J.   The Economic Theory of Cost of Living Index Numbers.
    New York: Columbia University Press.   1949.   pp. 106.   $2.00.
U. S. DEPARTMENT OF COMMERCE, BUREAU OF THE CENSUS.   Forecasts of the
    Population of the United States, 1945–1975.   Washington, D. C.: The
    Department.   1947.   pp. 113.   45 cents.
WALKER, GILBERT.   Road and Rail.   London, England: George Allen and
    Unwin, Ltd.   1947.   pp. 282.   $3.75.  [Distributed by The Macmillan
    Company in the U. S.]
WEBER, ADOLF.   Kurzgefaszte Volkswirtschaftslehre.   Berlin and München:
    Lieferschein.   1948.   pp. 251.   DM 8.60.
WILLIAMS, ELGIN.   The Animating Pursuits of Speculation.   New York: Columbia
    University Press.   1949.   pp. 230.   $3.25.
WILSON, THOMAS.   Fluctuations in Income and Employment.   New York:
    Pitman Publishing Corporation.   1949.   pp. 217.   $4.00.
WRIGHT, CHESTER W.   Economic History of the United States.   New York:
    McGraw-Hill Book Company.   1949.   pp. 941.   $6.00.
ZIPF, GEORGE KINGSLEY.   Human Behavior and the Principle of Least Effort.
    Cambridge 42, Mass.: Addison-Wesley Press, Inc.   1949.   pp. 573.

# Harvard Economic Studies

## New Books

## AGRICULTURE AND INDUSTRIALIZATION
### AN INQUIRY INTO THE ADJUSTMENTS THAT TAKE PLACE AS AN AGRICULTURAL COUNTRY IS INDUSTRIALIZED

*By Pei-Kang Chang* — With particular reference to China, the author brings the theory and the history of economics together in the first systematic study of the adjustments between agriculture and industry during the process of industrialization in an agricultural country.
*260 pages plus index, $5.00*

## THE ECONOMIC PROBLEMS OF FORESTRY IN THE APPALACHIAN REGION

*By William A. Duerr* — The whole array of problems found at this meeting ground of the biology, engineering, and economics of forestry are herein described and analyzed.
*xvi + 317 pages, $5.00*

## THE DEVELOPMENT OF THE AMERICAN GLASS INDUSTRY

*By Pearce Davis* — This is the first economic history of the development of the American glass industry from small-scale handicraft to full-blown mass production. It includes the first comprehensive study of the effects of American tariff policy and of organized labor upon that development.     *Coming June 22.*     *305 pages plus index, $6.00*

### NEWLY REPRINTED

## MONOPOLISTIC COMPETITION AND GENERAL EQUILIBRIUM THEORY

*By Robert Triffin* — The fourth printing of Mr. Triffin's book, well known for its thoughtful, compact, and provocative summary and synthesis of monopolistic competition doctrine, will be welcomed by teachers and students alike.     *xiii + 197 pages, $3.00*

## THE THEORY OF ECONOMIC DEVELOPMENT
### AN INQUIRY INTO PROFITS, CAPITAL, CREDIT, INTEREST, AND THE BUSINESS CYCLE

*By Joseph A. Schumpeter* — "Nowhere else does one find so penetrating a treatment as here of the significant characteristics of the process of development of economic dynamics, in contrast with the 'circular flow' of an essentially static society." — ALVIN H. HANSEN, *The Journal of Political Economy.*     *Third printing, xii + 255 pages, $4.00*

*The Harvard Economic Studies now comprises 86 volumes covering a wide variety of aspects of economic history and theory. A complete list will be furnished upon request.*

*At all bookstores, or from*

## HARVARD UNIVERSITY PRESS
### CAMBRIDGE 38, MASSACHUSETTS

# THE

# QUARTERLY JOURNAL

# OF ECONOMICS

## BASING POINT PRICING AND PUBLIC POLICY[1]

### SUMMARY

I. Introduction: the nature of the problem, 289; alternative pricing systems, 292; standards of public policy, 293. — II. Pricing practices and the level and flexibility of margins, 294; and the level of costs, 297. — III. Pricing practices and the long run adaptation of capacity to demand, 300; the location of producers, 304; and the allocation of transport resources, 305. — IV. Pricing practices and capacity in relation to cyclical fluctuations in demand, 306; and the organization of the industry, 309. — V. Summary and conclusions, 313.

The Supreme Court's reaffirmation (by a 4–4 tie vote without written opinion) of the Federal Trade Commission's decision in the Rigid Steel Conduit Case finally settled the legal status of basing point pricing systems.[2]  The earlier Cement Case[3] had defined any agreement to use basing point pricing as unlawful, and put a wide construction on what was evidence of an agreement. In the Conduit Case the Federal Trade Commission ruled that the concurrent use of basing point pricing by numerous competitors was sufficient evidence of unlawful conspiracy.  Further, it held that individual use of the system by a single firm in the knowledge that it was being used by competitors was likewise illegal.  The Circuit Court, in upholding the Commission, asserted that the legal question presented by that case was identical with that presented by the Cement Case.  Under no circumstances, therefore, could basing point price systems be free

1. The material in this paper was first presented to an informal discussion group on anti-trust policy consisting of Prof. M. A. Adelman of M. I. T. and Professors S. S. Alexander, R. R. Bowie, D. F. Cavers, L. Gordon and Dean E. S. Mason, all of Harvard. I am much obliged to them for valuable suggestions and corrections.  They are not, of course, either individually or collectively responsible for the argument here presented or the errors it may contain.

2. The FTC decision is given in Triangle Conduit and Cable, Inc., et al, FTC Docket 4452 (1948).  The decision was upheld on appeal by the 7th Circuit Court of Appeals, 168 Fed. 2d 175 (1948).  The high court decision is recorded in 1949.

3. FTC vs. Cement Institute, et al, 333 United States 683 (1948).

of the vice of unlawful conspiracy, which make them "unfair" methods of competition" under the Federal Trade Commission Act.

This decision will probably give further impetus to the discussion of legislation to make basing point pricing legal by amending the Clayton and Federal Trade Commission Acts. The discussion was initiated in the Eightieth Congress by the Republicans through the medium of a subcommittee of the Senate Committee on Interstate Commerce, under the chairmanship of Senator Capehart of Indiana (Rep.). It is being continued in the present (Eighty-first) Congress; and a bill was introduced to effect such amendments (S.236) by Senator Johnson of Colorado (Dem.) early in January.

The need for dispassionate examination of basing point pricing and the possible alternatives to it in terms of public policy is clear. The present essay is an attempt to do so, by no means the first in this knotty field.[4]

Before examining the tests by which public policy in this field should be measured, let us summarize briefly the characteristics of markets in which the basing point pricing system is used, and the chief features of the system in operation. The characteristics which follow are typical, but not every industry operating under the system will possess them all.[5] The product sold is essentially standardized, so that the output of one producer at a given consuming point is a perfect substitute for the output of another producer at that point. Therefore, in an equilibrium situation, the prices charged by two producers at any given point must be the same. The product also is low in value per unit weight, so that for shipments over all but very short distances, the transportation cost forms a substantial fraction of the delivered price. Thus spatial differentiation of the

4. Some of the recent articles have been: Frank A. Fetter, "Exit Basing Point Pricing," and Corwin D. Edwards; "Basing Point Decisions and Business Practices," both in *American Economic Review*, December, 1948; J. M. Clark, "Law and Economics of Basing Points," *American Economic Review*, March, 1949; and Edwin B. George, "The Law and Economics of Basing Points" (3 articles), *Dun's Review*, September, October, and November, 1948. A book on the subject was published in April, 1949: Fritz Machlup, *The Basing Point System*, (Philadelphia: Blakiston, 1949).

5. In general, factual material on the operation of the basing-point system is not too abundant, and nearly all that is available bears on only two industries of the many in which the system was used — steel and cement. All the factual material referred to in this paper is drawn from the experience of one or another of these industries. This experience is described in various places. See especially: A. R. Burns, *The Decline of Competition*, Ch. VI, "Price Discrimination"; *TNEC Hearings*, Parts 19, 20 and 27; TNEC Monographs 42 *The Basing Point Problem* and 41, *Price Discrimination in Steel;* and the Supreme Court opinion on the Cement Case, 333 U. S. 683 which provides a good brief summary of the operation of basing point pricing in the cement industry.

product, delivered to the consumer, forms an essential element of the system. Shipments do, in fact, take place over fairly long distances. This can arise either from economies of scale which make local producers serving local markets uneconomical, or from strong locational factors which lead to the concentration of production in a few areas, or both. The cement industry furnishes an example of the operation of the first reason; the maple flooring industry (located in the region of timber stands in Michigan and Wisconsin), of the second; and the steel industry, the classic example of basing point pricing, of the combination. The efficient scale of operations is large, capital investment in an economical plant is fairly great per unit of output, and the ratio of marginal cost to average cost is low for all rates of operation below "full capacity." The rate of operation of the plants in the industry is frequently below full capacity, often because there are large cyclical fluctuations in demand for the product. Production equipment employed is specialized and long-lived, and thus exit from the market, either by product substitution or by allowing plant to "die" through non-replacement, is a difficult and protracted process. The market demand for the product is generally inelastic at and below prices which correspond to output considerably less than the full capacity of the industry. Finally, the market contains few enough sellers so that oligopolistic calculation plays an important role in the actions of firms in the market. This is clear enough in the steel industry. The large number of firms in the cement industry seems to contradict this statement, but shipments of cement rarely move more than 200 or 300 miles and in any local market the number of competing firms is very small, fewer than 10 in most areas. The combination of small numbers and spatial differentiation makes these industries classic examples of monopolistic competition.

The mechanics of basing point pricing are well known, and need not be rehearsed here. Certain features of the operation of the system, however, deserve the emphasis of repetition. Its outstanding feature is the creation of a fixed, well-defined price structure, with delivered prices of all sellers identical to all consumers at each specific location. The effective operation of this system requires basing points and base prices publicly known (in the trade), and uniquely defined freight costs from every basing point to every possible consuming point. This last requirement is usually fulfilled by the use of a common compilation of freight rates in the form of a freight book by all firms in the industry. In practice, the system usually operates on the basis of all-rail freight in the calculation of delivered prices, although it could conceivably forego this restriction. Since rail

freight costs must be paid, rail freight is used almost to the exclusion of water and truck transport.[6]   Lastly, there is always a significant amount of market interpenetration, defined as occurring when mills sell to customers located where mill prices plus freight costs from other mills which could handle their orders are lower than from the selling mill.   This market interpenetration usually involves freight absorption (when there is a multiple basing point system) and always involves selling costs in excess of what would be spent without market interpenetration.   The end result of market interpenetration is a complex structure of geographical price discrimination, determined by the locational pattern of mills and consumers, and the actual degree of market interpenetration.

So much for the circumstances and consequences of basing point pricing.   In order to examine this pricing practice from the point of view of public policy recommendations, it is necessary to formulate both a set of alternative situations which are feasible of attainment through governmental action, and a set of standards by which the desirability of the alternatives can be tested.   There are four models which cover fairly well the essential features of the wide range of geographic pricing practices:

(1) *Single basing point model.*   The use of a single basing point; or a single major basing point and subsidiary basing points with base price differentials over the primary base of the same order of magnitude as the costs of shipment between them.   Basing point quotations are strictly adhered to; all freight computations are made on an all-rail basis.

(2) *Universal freight equalization model.*   Every producing mill is a basing point.   There is no f.o.b. mill selling, and standardized all-rail freights are used in calculating delivered prices.   In this model, the relationship of the base prices of the several producers can range from rigid price leadership, in which differentials among the various mills always remain the same, to complete independence in setting base prices.

(3) *Uniform f.o.b. mill price model.*   Each seller maintains an announced mill price which is the same to all buyers (at any given time).   The buyer takes possession at the mill, and chooses the method of delivery and pays the freight.   Here again, the model does

6. In steel, consumers were allowed to take delivery in their own trucks on payment of 35 per cent of rail transportation costs.   See *TNEC Hearings*, Part 27, pp. 14182 ff.   See also Part 20, pp. 10830–45, and correspondence on pp. 11005–6 for the reaction of the Steel Corporation to attempts of a consumer to take delivery in his own barge at a point not the point of consumption.

not specify the relationship between the mill prices of the various sellers.

(4) *F.o.b. mill selling with price discrimination model.* The sellers maintain no announced prices, but deal with each customer as best they can. As in (3), customers can choose the method of transportation and pay the freight. When base prices are independently determined, model (2), universal freight equalization, (if modified by allowing the customer the election of delivery method) amounts substantially to model (4).

These models differ only in the specification of pricing practices; it is assumed throughout that the basic features of the industries using basing point pricing remain unchanged, except insofar as the change in pricing practice in itself reacts on the structure of the industry and changes it. This is, of course, the assumption appropriate to an examination of the effects of possible changes in pricing formulae on the functioning of industrial markets. It is assumed further that this range of pricing practices represents possibilities which could be achieved in practice by government action, an assumption for which there is some justification. The first model was more or less exemplified in the operation of Pittsburgh Plus pricing in the steel industry until the U. S. Steel Corporation abandoned it under pressure from the Federal Trade Commission in 1924. A stage between the first and second was represented by the recent history of the steel and cement industries, until the Cement Institute decision. The third model is what would seem to be the only legal method of pricing in the present state of the law. Given a change in the interpretation of the Robinson-Patman Act shown in recent Court decisions[7], or its outright repeal, the fourth model might be taken as a rough prediction of how industry pricing practices could respond to the law on basing point pricing as laid down in the Cement and Rigid Steel Conduit Cases.

The proposed standards for evaluating the alternative pricing practices are likewise four in number.

(1) What is the effect of the pricing system on the level and rigidity of prices? This involves the effects of the pricing system on the flexibility of margins, the ease with which downward adjustments

7. Such as *United States v. Great Atlantic and Pacific Tea Company*, CCA 7th, 1949; and *Standard Oil Company v. Federal Trade Commission*, CCA 7th, 1949. On this whole matter see the stimulating discussion by M. A. Adelman, "Integration and Antitrust Policy," to be published in the *Harvard Law Review*, October, 1949 and his article "The A & P Case: A Study in Applied Economic Theory," this *Journal*, May, 1949,

in margins can occur, and the effects of pricing practices on the level of costs.

(2) Do the pricing practices facilitate or retard the adjustment of capacity to demand in the long run? This question has two aspects — the problem of aggregate adjustment, and the problem of the regional and local balance between capacity and demand. The second aspect shades over into the problem of the influence of pricing practices on the general locational structure of industry.

(3) What effect does the pricing system have on cyclical fluctuations in prices and output? Since the basic characteristics of the industries in question guarantee that their capacity cannot be adapted to cyclical fluctuations in demand, this question really asks how well or how badly the pricing system functions in helping the industry to achieve the appropriate minimum average rate of return over the cycle.

(4) How does the pricing system affect the organization of the industry? To what extent are the number of producers, the scale of production, and the relative viability of large scale and "independent" producers in the competitive struggle affected by pricing practices?

The application of these four tests to the four models will give each model a set of "marks"; a comparison of these marks should give some guidance on the desirable aims of government action.

## II

In considering the relation between pricing formulae and the magnitude and flexibility of margins, the basic oligopolistic character of the markets under consideration must again be stressed. In a situation of unused capacity and inelastic demand, what prevents "cutthroat competition" from driving price down to marginal cost? It is oligopolistic rationality — the realization by each seller that his cuts will be followed by his rivals, and that therefore he has little to gain by cutting, and much to lose by initiating a process of reducing margins. The importance of pricing formulae in this situation lies in the extent to which they make it easy for each member of the group to follow what every other member of the group is doing. For example (in addition to the important role it has in making possible an intelligent comparison of the offers of various sellers by buyers in a market with an extremely complex product structure) the "extra list" in steel serves a vital function in reducing the dimensionality of the price structure, and thus making it feasible for the rival sellers to compare each others' price quotations. And it is significant that

the extra list is explicitly agreed upon among the various producers.[8] In an oligopolistic market, especially one such as steel in which sales are frequently made on contracts for large quantities running over periods of time, it is almost inconceivable that price reductions would be made if they had to be made uniformly to all comers at one fell swoop. Rather some kind of piecemeal reduction, involving various kinds of price discrimination, is to be expected; and, in fact, seems to be the rule. Any reduction of quoted prices is usually preceded by a period of "price shading," during which "secret" concessions are made to customers, concessions which differ widely as between customers.[9] In this process of discriminatory price reduction, various asymmetries in the market undoubtedly play an important role: large customers get better prices than small ones; smaller producers, or financially weak producers, may be the first to offer concessions; producers located in relatively disadvantageous positions may be prompted to strive, through concessions, for more business in local areas where mill nets are high; and so forth.

To the extent that a pricing system formalizes quotations, and publicizes (in the trade) the "proper" prices at which transactions should be made, it discourages the processes by which prices are reduced. Judged on this ground, the single basing point system is undoubtedly the worst offender among our four models. The only "proper" channel for price change is controlled entirely by the single producer who sets base prices. He is formally the price leader for the industry, and knows that any base price changes he makes are changes in the industry's prices, not only in his own. Moreover, the fact that under this system, the basing point producer can penetrate into any market without suffering lowered mill nets might well make for greater caution on the part of other producers in making secret concessions. A universal freight equalization system may in fact function as a rigid system of price leadership, and be just as effective in narrowing the amplitude of price adjustments as a single basing point system. But the very fact that many producers can, at least potentially, set base prices independently, and the further fact that limits to market interpenetration, and thus to "punitive" invasions of markets, are set by the costs of freight absorption, probably make

8. See *TNEC Hearings*, Part 19, pp. 10557–80, testimony of Benjamin Fairless and other United States Steel executives, and pp. 10621–35, testimony of Eugene G. Grace, president of Bethlehem Steel Company.

9. See TNEC Monograph 41, *Price Discrimination in Steel*, for evidence of discrimination between large and small purchasers of steel in a period of slack demand. Also, see Table I below for an indication of the range of price concessions in steel.

such a system more flexible in its operation than the single basing point system.

It seems doubtful that uniform f.o.b. mill pricing would produce any better results from the point of view of price flexibility than a universal freight equalization system, and it might even produce somewhat inferior ones. The uniform, published mill price of each producer would certainly make each seller's price policy highly visible to his rivals. Each seller would fix his attention on the market area of his rival, rather than on his delivered prices, as he would under any basing point scheme; but the rival's price behavior would be revealed just as clearly. Moreover, although each seller would be free to fix his own mill price (provided he wanted to exercise such freedom in an oligopolistic market) any change in price he made would have to be uniform to all customers. This enforcement of uniformity, added to the "visibility" of his behavior to rival oligopolists, would do much to discourage any producer from initiating a price change. Finally, uniform f.o.b. mill selling would end market interpenetration. Market interpenetration in the context of a rigidly operating basing point system with strong price leadership, has no competitive virtues. However, in a situation characterized by the possibility of discriminatory price shading, market interpenetration, which spreads the rivalry among sellers over a large area of the market, instead of concentrating it along the fringes where the market areas of rival sellers meet, undoubtedly increases the speed of price changes and probably their magnitude also.

The foregoing argument shows that the fourth model, f.o.b. mill selling with price discrimination allowed, would be the one in which the pricing method would offer the least assistance to rival sellers in behaving "rationally" in accordance with the oligopolistic character of the market, and would discourage the formation of a fixed pattern of price leadership. The question arises as to whether such a situation would not lead to "too much competition," with prices cut to marginal costs whenever unused capacity existed, and even efficient firms driven into bankruptcy.[1] No categorical answer can be given to this question. It is the author's opinion that the danger is unlikely, for although the abolition of formula pricing would make it more difficult for firms to abide by the oligopolistic calculations which militate against price reduction, it would not eliminate such calculations entirely. The fundamental oligopolistic character of the market appears to be a sufficient brake on really destructive competition.

Thus, in terms of the four models of pricing systems, the greatest

1. See Professor Clark's article, *op. cit.*

gain in the flexibility of margins is promised by a change from a single basing point system, or a multiple basing point system with fairly rigid price leadership, to f.o.b. mill pricing with discrimination. A change to uniform f.o.b. mill pricing seems to offer little along these lines; a little more may be expected from independent base pricing in a multiple basing point system; i.e., the elimination of price leadership.

In practice, in recent years basing point pricing systems have been of the multiple basing point variety, with fairly strong price leadership. The increase in flexibility of margins which can be expected from changes in the direction of more independence in pricing depend on the rigidity with which the systems have operated in the past. Some light is thrown on this point in a study by the Bureau of Labor Statistics of the prices actually paid for steel products by a large sample of consumers over the period 1939–42. Some of the results of this study are given in summary form in Table I below. This tabulation suggests that there is still room for a substantial increase in the magnitude of price reductions in periods of low output. In the second quarter of 1939, when output ran about half of rated ingot capacity, average price concessions ranged from eight per cent in sheets to only three per cent in structural shapes. These averages concealed wide variations as between customers, some of whom paid prices as much as 25 per cent under published prices. But these were only a small proportion. A year later, with output at 72 per cent of rated ingot capacity, the spread of concessions was much reduced, even though average prices increased but little. Thus, in hot rolled sheets — the product for which concessions were greatest — a quarter of the sales were made at concessions of 15 to 25 per cent on the published price in 1939. By the following year, the number of sales at prices 15 per cent or more under the published price had fallen to fewer than five per cent of the total. All in all, the figures of Table I speak well for the efficacy of the basing point system as a method of price leadership in the steel industry.

The level of prices depends on the level of costs, as well as on the seller's margin. To the extent that different pricing systems affect differently the level of costs, the choice between them may influence the level of prices. Any pricing system which allows market interpenetration adds the cost of freight absorption to the other costs of the mill; it is the mill net yield and not the delivered price less the calculated freight which is the "price" from the point of view of the seller. This cost is present in three of the four models; only with uniform f.o.b. mill selling is market interpenetration ruled out

TABLE I

ACTUAL PRICES COMPARED WITH PUBLISHED PRICES[1]

| Product and Year (Second Quarter Figures)[2] | Operating Rate (Per Cent Ingot Capacity) | Per Cent of Number of Sales made at Actual Prices which were given Percentages of Published Delivered Price[3] | | | | | | | Average Ratio of Actual to Published Delivered Price[3] |
|---|---|---|---|---|---|---|---|---|---|
| | | 75 to 80 | 80 to 85 | 85 to 90 | 90 to 95 | 95 to 98 | 98 to 101 | over 101 | |
| **Hot Rolled Sheets** | | | | | | | | | |
| 1939 | 51 | 5.2 | 20.4 | 12.4 | 19.7 | 19.7 | 11.0 | 11.6 | 92 |
| 1940 | 72 | 1.5 | 2.9 | 7.1 | 60.0 | 4.3 | 24.2 | .. | 94 |
| 1942 | 98 | .. | .. | .. | 0.4 | 1.2 | 96.0 | 2.4 | 101 |
| **Cold Rolled Sheets** | | | | | | | | | |
| 1939 | 51 | .. | .. | 28.6 | 28.5 | 7.1 | 32.2 | 3.6 | 95 |
| 1940 | 72 | .. | 2.1 | .. | 74.6 | 4.2 | 19.1 | .. | 95 |
| 1942 | 98 | .. | .. | .. | .. | 3.2 | 93.6 | 3.2 | 101 |
| **Hot Rolled Strip** | | | | | | | | | |
| 1939 | 51 | 16.2 | 12.9 | 9.6 | 9.6 | 22.6 | 17.2 | 12.9 | 92 |
| 1940 | 72 | .. | 2.0 | .. | 59.3 | 8.1 | 26.6 | 2.0 | 95 |
| 1942 | 98 | .. | .. | .. | 1.7 | .. | 98.3 | .. | 100 |
| **Plates, Universal and Sheared** | | | | | | | | | |
| 1939 | 51 | .. | 3.0 | 6.0 | 9.0 | 12.6 | 61.0 | 8.4 | 97 |
| 1940 | 72 | .. | 0.9 | 2.5 | 8.8 | 12.5 | 70.2 | 5.1 | 98 |
| 1942 | 98 | .. | .. | .. | .. | .. | 85.0 | 15.0 | 103 |
| **Structural Shapes** | | | | | | | | | |
| 1939 | 51 | .. | 1.0 | 8.2 | 16.5 | 10.4 | 61.8 | 2.1 | 97 |
| 1940 | 72 | .. | 1.0 | 1.6 | 15.8 | 13.5 | 67.0 | 1.3 | 98 |
| 1942 | 98 | .. | .. | .. | .. | .. | 96.6 | 3.4 | 101 |

1. The data presented in this table come from a study made by the Bureau of Labor Statistics in 1943 entitled *Consumers' Prices of Steel Products*. The study was prepared for the use of the Office of Price Administration and the War Production Board. It was reprinted in *Iron Age*, April 25, 1946, under the title "Labor Dept. Examines Consumers' Prices of Steel Products." The study was based on an examination of data furnished by 629 companies consuming steel products, widely distributed by region and industry. The sample included only carload lot purchasers, and excluded distributors, warehousers, and subsidiaries of steel-producing companies. The companies in the sample accounted for 15 per cent of steel consumption in 1940. For each company, price information was obtained only on products purchased more or less regularly (monthly, if possible) in carload lots (except for alloys) between 1939 and 1942, and for which customer specifications remained constant over the period. The very largest consumers (e.g., the major automobile producers) are excluded from the sample, and therefore the actual extent of concessions is probably understated somewhat. The study presents data for the third quarter, 1939 and the second and fourth quarters, 1941 as well as for the three quarters shown in Table 1. In addition to the products for which prices are shown there, the study gives figures for cold rolled strip, merchant bars, and cold finished strip. The original tabulations are made in items of class intervals of one per cent (ratio of actual price to delivered price) which are here consolidated for brevity. Note that in every case the largest concession is included. The selection of time periods and products given in Table 1 is, in the author's opinion, a sufficiently representative one to yield a fair picture of the results of the whole study. Neither for the other products or for other time periods does the pattern of actual prices relative to quoted prices and the utilization of capacity appear significantly different from that here presented.

2. The percentages given are for the second quarters of each of the years 1939, 1940 and 1942.

3. The published delivered price was built up for each sale from the appropriate published base price, the freight from the applicable basing point, and the published extras. The published delivered prices entering into the comparisons of actual and published prices are those of April 1942. But these differed if at all from published prices in the various quarters only by very small amounts, since extras remained constant for the period, the greatest change in base prices was a few per cent, and freight rates increased about six per cent in March 1942. The error introduced by the use of 1942 prices runs in the direction of exaggerating the magnitude of the concessions in the second quarters of 1939 and 1940: the size of this error is very small.

entirely. Some notion of the magnitude of the cost of freight absorption can be gained from a study of steel shipments made for the TNEC.[2] The study covered the month of February, 1939, a period of slack demand in which it is likely that market interpenetration was extensive. Freight absorption in this month ranged from about three to five per cent of the delivered price for various products. This is a small cost; nor is it certain that it is all reflected in price. In an oligopolistic market, part of it may be absorbed by sellers through a reduction in profits.

A rigid basing point system with price leadership eliminates price competition and channels rivalry into selling efforts. Since products sold under basing point pricing are characteristically standardized, the cost of selling effort is largely remuneration of salesmen. Such selling costs will be present to some extent under any pricing system in which market interpenetration exists and customers must choose between the products of rival sellers offered at the same delivered price. This will obviously be the case under any form of basing point system. Under f.o.b. mill selling with price discrimination, there will be occasions on which a seller offers to meet rather than shade a rival's price, and relies on a combination of selling efforts and customer loyalty to swing the buyer's decision. Under uniform f.o.b. mill selling, pure selling costs would theoretically disappear, if mills were widely separated and the freight rate structure such that boundaries between mills were lines rather than regions. In fact, neither of these conditions is met. In many cases rival mills are located close to each other so that there are numerous consuming locations at which prices from two sellers would be the same.[3] Moreover, the existence of "blanket rates" covering long distance shipments to and from large areas would produce overlapping market areas under f.o.b. mill pricing even if rival sellers were not located side by side. In any case, the total cost of the sales force cannot be considered as an "unnecessary" selling cost attributable to the existence of non-price competition. The salesmen perform certain

2. See *TNEC Hearings*, Part 27, appendix, pp. 14331–14428. The study was based on a sample of steel shipments of all major producers in the month of February, 1939. The sample covered shipments aggregating some 600,000 tons, or about 25 per cent of the total shipments of steel during the month. February, 1939 was a period of slack operations, about 50 to 55 per cent of rated ingot capacity.

3. See, for example, the map of counties in Western Pennsylvania and Eastern Ohio, showing towns in the market areas of various steel sheet mills under an f.o.b. mill selling system with an assumed level of mill prices; *TNEC Hearings*, Part 27, appendix, p. 13832. This map was prepared by the United States Steel Corporation.

necessary functions in the distribution process; at a minimum, order taking, and where the product is complex, as in steel, the transmission to buyers of information on specifications and the suitability of the many grades and types of product for various uses. To the extent that salesmen's pure "selling" functions are performed jointly with these other duties, no extra cost of selling exists. All in all, it is doubtful that changes in pricing practices will produce significant changes in selling costs through lessening the importance of non-price relative to price competition.

The possibility of cost saving on the freight bill through the use of cheaper means of transport under other than basing point pricing, will be discussed below in considering the relation of pricing practices to location and utilization of capacity.

Finally, the influence of flexibility of margins itself on costs should not be neglected. The existence of fairly strong downward pressure on margins in periods of unused capacity, which would be expected under some pricing arrangements would be a strong stimulus to energetic efforts at cost reduction by management. The use of the most efficient known production techniques is not, in practice, something that occurs automatically. Every change in methods is painful (and even costly) in a large organization, and the existence of a fairly stable profit margin per unit tends to favor let-well-enough-alone policies. This aspect of the relation of prices to costs will probably be more important, in the long run, than the (at least conceptually) more easily measured differences in freight absorption and selling costs contingent on the use of differing pricing practices. It is another important argument in favor of giving high marks to the fourth model, or pricing practices resembling it.

## III

The second standard for the evaluation of alternative pricing systems is the relation between the pricing system and the adaptation of capacity to demand in the long run. There is no general theoretical rule on the responses of an oligopolistic group of suppliers to long run changes in demand. The persistence over time of high margins (which formula pricing facilitates) would indicate some restriction of supply in comparison with what would happen in an otherwise similar competitive market. This is not a very helpful criterion in practice, however, and something more concrete is needed. The impact of the pricing system on the adjustment of supply depends on the direction of change in demand over time; in particular, the case of an increase

in demand over the long run must be treated separately from the case of stationary or declining demand.

If demand is stable in the long run and excess capacity (the existence of which is postulated as characteristic) is present only during cyclical troughs, capacity may be considered as adjusted to the long run level of demand. The problem of cyclical adjustment which remains in this case will be treated in the following section. The situation in which excess capacity exists at cyclical peaks as well as during troughs with the long run level of demand stable is similar to one in which the long run level of demand is declining, in that the appropriate adjustment requires the elimination of some producing units. Whatever adjustment takes place must be brought about by losses which are large enough and continuous enough to drive firms into bankruptcy, or at least to threaten them with it. It is true, of course, that bankruptcies may not be sufficient to bring about an immediate adjustment of capacity to demand, since they can result in the reorganization of old firms, or the creation of new firms in which the equipment of the original firm is valued at less than its reproduction cost. These successor firms may survive for a while; but the necessity of replacing old equipment will finally confront them with the same problems which caused their predecessors to fail. Again, the process of bankruptcy and reorganization might well be repeated in a new cycle, or even several new cycles. It must be assumed, however, that eventually old equipment will wear out, and potential investors and entrepreneurs will learn something from the previous history of the industry, and thus reductions in capacity will ultimately occur.

In a situation in which a reduction in productive facilities is required, it is clear that a working basing point system interferes with the proper adjustment. In the first place, to the extent that it functions as a device for maintaining high margins, the whole process of failure is slowed down. Moreover, a basing point system entails market interpenetration, which means that existing markets are shared among many producing points. The nearer a mill approaches failure, the greater would be its incentive to attempt to cut into any market, no matter how distant, in which sales promised some mill net return over marginal cost. Thus the impact of a decline in demand, which in general will be greater in some areas than in others, will not be concentrated on the mills located in the areas of most rapid decline, but spread out more or less evenly over all producers. This will further slow down the process of failure, as compared with what would happen if the mills in areas of more rapidly falling demand

sold only in their home markets. The spreading of the impact over all producers also means that differences in financial strength have relatively more, and differences in productive efficiency and location, relatively less influence in determining which firms will leave the industry than they would under a pricing arrangement which did not permit market interpenetration.

The hindrances to appropriate adaptation of output discussed above, except for the first — maintenance of high margins — arise out of market interpenetration, and their importance is measured by the extent of it. Market interpenetration is a feature of three of the four models; only under uniform f.o.b. mill pricing would it be absent. The extent of market interpenetration depends, in the other three models, more on the geographical pattern of demand, costs, and freight rates, than on the specific pricing practices in which the models differ. There is one exception to this, however; market interpenetration probably is greatest, other things being equal, under a single basing point system, in respect of the relations between the base mill (or mills) and all other mills. Thus in the special circumstances in which the home market of the base mill is the region of the greatest decline in demand, the single basing point system would lead to a slower reduction of capacity than a universal freight equalization or discriminatory f.o.b. mill system of pricing.

Demand may be rising in the long run (instead of falling or remaining stationary) and the appropriate adjustment of supply to demand will require additions to productive facilities. In this situation, excess capacity may be present at cyclical peaks as well as during troughs, because of a certain rate of building in anticipation of increasing demand, or it may exist only during cyclical lows. In either event, the effect of the pricing system on the expansion of capacity will depend on the geographic pattern of the change in demand. If it is fairly uniform from market area to market area, the existence of market interpenetration will not alter the local incidence of the stimulus to new investment. The only significant difference between the alternative pricing systems then which would affect the rate of investment would be the difference in profit margins expected under them. The higher margins expected under the more rigid formula pricing arrangements might then act to stimulate a more rapid increase in capacity than would be made with more flexible and competitive pricing systems.[4] This additional stimulus, however, might be cancelled in large part, or even overbalanced by

4. This argument is essentially similar to Professor Schumpeter's arguments in favor of monopoly. See *Capitalism, Socialism and Democracy*, Ch. VIII.

the greater pressure on the individual firm under more competitive pricing systems.

If, as is more likely, the growth of demand is geographically uneven, market interpenetration becomes more important. This is most clearly seen in a situation in which the expansion of demand takes place entirely in new regions, not near existing mills, while demand in old regions remains stable or even declines. Under uniform f.o.b. mill pricing, a new mill would be constructed in the region of new demand whenever entrepreneurs and investors foresaw enough demand in the region to maintain an economically-sized unit in operation at an average output rate which would yield a sufficient profit. Under the other three pricing systems, characterized by market interpenetration, the entrepreneurs in the new region would have to anticipate penetration of their market by sellers in the old regions. Nor could they expect to retaliate by penetrating markets in the old regions to the same extent, since these regions would be crowded with suppliers, and their long standing relationships with local customers would be a further barrier to new sellers. Thus the building of a new plant in the new region would wait for a higher level of demand under pricing systems characterized by market interpenetration than under a uniform f.o.b. mill system. The same kind of forces would be at work, although with diminished strength, in a situation in which there were large differences in the rates of change in demand among several regions without differences in the direction of the change. Thus the use of any pricing system which permits market interpenetration will lead to a slower rate of increase of capacity in response to a given rate of growth of demand unevenly distributed among regions, than would take place under uniform f.o.b. mill pricing.

This lag in the regional adjustment of capacity to demand is not undesirable. In many situations it may provide a closer approximation to an "ideal" utilization of resources than would a more prompt increase of capacity. The general economic rule for the substitution of new for old (but still useful) capital equipment is that such replacement is desirable only when average costs with the new equipment are less than marginal cost with the old. Under the conditions analyzed above, with demand increasing rapidly in a new region, but remaining stable in an old region where there is excess capacity, it is economic to supply the new region from the old plants as long as the marginal cost of production in the old plants plus transportation to the new region is less than the average cost of production would be for a new plant in the new region. Under uniform f.o.b. mill pricing,

as argued above, new plant would be built in the new regions while there was still unused capacity in the old regions. Since marginal costs are characteristically much below average costs at outputs below capacity in the industries under consideration, this can mean that the new plants represent uneconomic additions to capital. With market interpenetration, the addition of new capacity will wait until much more of the old capacity is utilized. The reasons for not anticipating a general lowering of prices to the level of marginal costs under f.o.b. mill pricing when unused capacity exists such as would bring the new areas within the market areas of the old mills have been examined above (page 296).

The appropriate adaptation of capacity to demand implies an economical geographic distribution of capacity, given the pattern of demand, as well as a suitable aggregate volume of productive facilities. Therefore, the possible impact of alternative pricing systems on the location of producers must be examined. And, since the location of producers of products which are further fabricated may affect in turn the location of the fabricators, an examination of their location must also be included. In general, the location of producers depends on the relations between the locations of raw materials, labor supplies, markets, and transportation costs. For a pricing system to influence the location of producers, it must work through one of these factors. In cement, the location of the consumers is fixed independently of the location of producers; and of the two major raw materials, coal and limestone, the more important limestone is nearly ubiquitous. In these circumstances the use of one rather than another pricing system would have no locational influence on producers, and could not, in the nature of the case, have any on consumers.

In steel, there is a more complex situation. Producers of steel are heavily raw material oriented; they locate in points where the assembly costs of coal, ore and subsidiary raw materials are low; but markets are not without locational pull.[5] Markets consist in fabricators of steel; many of these in turn locate with reference to the source of their major raw material. To the extent that a pricing system makes the geographical pattern of steel prices different from the "real cost" pattern, consumers will locate in accordance with the geographic price structure rather than with the underlying real cost structure. The location of consumers may in turn influence the location of expansion in steel production, which further reacts on consumers, and so on. Thus, a cumulative process of distortion of

5. See W. Capron and W. Isard, "The Future Locational Pattern of Iron and Steel Production in the U. S.," *Journal of Political Economy*, April 1949.

locational patterns can occur. This process is not infinite, since the fabricators have to consider the costs of reaching out to their own markets which increase as they all concentrate at one point. This process is best exemplified under single basing point pricing. Under this system, locations at the basing point are superior to locations at non-basing point mills for the fabricators, other things being equal. Therefore, fabricators tend to concentrate near the basing point, and mills there tend to grow relative to mills at non-basing point locations. This is true even though the basing point ceases to be the lowest cost production point, as it may well have been at one time. Thus, under Pittsburgh Plus, the rate of expansion of steel production in Chicago and Birmingham relative to Pittsburgh[6] was probably slowed down substantially.[7]

The locational pull of base mills operates only relative to non-base mills. Under universal freight equalization, or any other system in which each mill is in effect a base, it would disappear. Nor is there any difference between uniform f.o.b. mill pricing and the other two pricing models in this respect: market interpenetration in itself has no localizing effects.

The alternative pricing systems here discussed may have a further impact on location through their effect on transportation costs. Basing point systems, both single and multiple, in which prices are calculated on the basis of all-rail freight discourage the use of other methods of transportation by the consumer. This means that in steel, for example, a potentially advantageous site connected to a supplier by water will not have its transportation advantages considered by a fabricator who is choosing a plant location, since he cannot reap the savings arising from this location. The same is not true, of course, of the steel producer. To the extent that his customers will accept shipment by water, he will ship by water and increase his mill net yield by the difference between rail and water freights, and thus he may consider the advantages of waterside location.

In addition to its possible indirect effects on location of production, the discrimination against cheaper forms of transport embodied in basing point pricing — models (1) and (2) — operates directly to distort the allocation of resources employed in transportation. Any alternative pricing system which gives the customer the option on the method of delivery, without taxing particular modes of transport,

6. See A. R. Burns "The Decline of Competition" pp. 340–45, esp. Figure 48.
7. The further effect of basing point pricing on the location of fabricators of steel producing goods to which fabrication-in-transit (f.i.t.) freight rates apply are not considered here. This is a problem more germane to a discussion of freight rates than to a general survey of basing point pricing.

will thus result in a more economical utilization of transportation resources. No indication of the magnitude of these economies can be given because of the lack of basic statistical data.

It is important to note that whatever advantages of efficiency in geographical allocation of resources the foregoing analysis indicates will result from the use of one rather than another pricing system, will apply only to plants as yet unbuilt. Thus the benefits of a change would be spread out over a long future period. On the other hand, some existing plants might become unprofitable as a consequence of a change from one to another pricing system, but little can be said in general terms about the magnitude of this transition problem.

## IV

The two remaining standards for grading alternative pricing arrangements are the effects of the pricing arrangements on cyclical fluctuations in output, and on the organization of industries using them. There is less to be said under these heads than under the two preceding, not because they are intrinsically less important, but rather because the questions raised under them are less amenable to general analytical answers than those previously discussed.

The essence of the cyclical problem in industries of the sort under discussion lies in the cyclical unadaptability of capacity to demand in the short run on the one hand, and the high income elasticity and low price elasticity of demand for their products on the other. This means that a capacity just sufficient to meet the minimum level of demand at the cyclical trough would fall far short of producing what was required (at constant prices) in a boom. Further, the increase in prices and profits caused by the impact of sharply rising demand on limited capacity would not lead to an increase in capacity within the short period. Conversely, capacity sufficient to meet boom-time peaks in demand implies the existence of much unused capacity during times of slack business, since price cuts would not greatly increase demand. The problem then is, how shall the costs of providing the peak load capacity be met.

Ideally, a two-part price system would answer the problem — marginal cost pricing for output, with any deficits separately financed from tax revenues, and the amount of plant decided by a central authority on the basis of the usual welfare criteria.[8] A fairly close approximation to this ideal method in practice could be achieved as

8. See H. Hotelling, "The General Welfare in Relation to Taxation and to Railway and Public Utility Rates," *Econometrica*, July, 1938.

follows. In any industry under consideration, private firms would own and operate enough plant to produce for periods of slack demand. Additional plant would be owned by the government, and turned over without fee to the private firms for operation during periods of peak demand. Marginal cost pricing would be prescribed. The amount of government plant, and the amount to be brought into production at any time would be determined by the prescription of some average normal profit rule for the private firms. While such a scheme might be practicable from an economic point of view, it would involve substantial administrative problems, and also, probably cannot be considered practicable politically.

Within the framework of purely private operation of industry, two general approaches to the problem of paying for the appropriate level of capacity over the cycle are possible. The first, represented by Professor Machlup in his recent book[9], is that the proper pricing rule requires marginal cost pricing at all times; in other words, pure competitive pricing. Under this rule, capacity would be determined at a level such that the returns during periods of high demand would cover the deficits incurred in periods of slack demand and provide for normal profits. Prices would fluctuate quite widely over the business cycle, since at any output level below full capacity, marginal cost would be much below average cost, while in the neighborhood of capacity, marginal cost would rise very sharply over a short range of output.

While this price policy — if it were followed — might be desirable in certain industries using basing point pricing, it is doubtful that it is suitable for steel and other basic industries producing investment goods. The demand for such products varies greatly over the business cycle. In general, demand for these products is price inelastic; this is especially true in periods of low and falling income, when investment is unresponsive to current cost changes in the face of poor expectations. An exception to the general rule may occur during the course of a recovery, especially after the most obvious and profitable investment opportunities have been exploited. At that time, a sharp rise in the price of investment goods might well react unfavorably on the volume of investment, and so shorten the recovery. Thus a steel industry which had capacity substantially short of that needed to meet peak demand in terms of a fixed price level (say, the average level over the cycle) might be an agent in cutting down the total level of investment and income over the cycle. This, in turn, would lead to further losses in the steel industry, and thus to further shrink-

9. *Op. cit.*, Ch. 6, esp. pp. 210–11.

age in capacity. Even if the initial situation were one in which capacity was larger than needed to meet peak demands, rather than the reverse, it is not certain that a flexible price policy would lead to a desirable adjustment. If capacity could change only by fairly large steps — equivalent to the output produced by a single firm — the first reaction to the initial situation could lead to a position of under-capacity, and further adjustments would again lead away from equilibrium instead of toward it. In general terms, it may be stated that a flexible price policy in an industry of the kind under consider-ation will lead to an equilibrium in capacity only if the cyclical fluctuations in aggregate income are independent of the price fluctua-tions in the products of the industry. Where it is unlikely that this is the case, as in steel, which is needed for many investment goods and has no effective substitutes over a wide range of uses, it is clear that there is no strong prima facie case for a cyclically flexible price policy.

The other approach to the question of cyclical price flexibility is that which involves a fairly great degree of stability, with prices substantially above marginal costs in periods of low output, and perhaps below marginal costs at the highest peaks of demand. There is no single, definite pricing rule which characterizes this view; it includes the range from the advocacy of strict full cost pricing, to the broad proposition that it is undesirable to allow prices to fall all the way to marginal costs during slumps in demand. Something near the first view is typically the view of businessmen in the industries under consideration; many economists, of whom J. M. Clark is one, have expressed the second.[1] Acceptance of the general proposition that complete flexibility in the sense discussed above is not desirable, leaves unanswered the question of how much inflexibility is necessary, and what arrangements will produce it. Industry advocates of basing point pricing advance this system as a necessary barrier to undesirable price flexibility. The validity of this contention turns on the factual issues of profitability and the adequacy of investment. It is difficult to examine these factual issues in detail, because the necessary infor-mation is lacking. Nevertheless, in the important case of the steel industry, the level of profits (even during the thirties) and of invest-

---

1. For an expression of the business view, see the now famous letter of John Treanor of the Cement Institute, cited in Machlup, *op. cit.* p. 41 n. in which the cement industry is characterized as one which "above all cannot . . . stand free competition." It is clear from the context that price competition is meant. For an expression of Clark's views see the article in *American Economic Review, op. cit.*, and "Basing Point Methods of Quoting Prices," *Canadian Journal of Economics and Political Science,* November, 1938.

ment do not suggest that the multiple basing point system just barely succeeded in maintaining the necessary capacity. The analytical argument stated above in connection with the discussion of the level of margins and prices (page 294) applies here too; given the fundamentally oligopolistic character of the markets in question, the danger of extreme price competition which would drive prices down to marginal costs in periods of slack demand seems small. The recognition by each of the rival sellers in the market of the relations between his price policy and those of the others should suffice to restrain the downward movements of prices in cyclical troughs before they reach danger levels. This presumes that the differences in costs among the firms are not so great that a price which seems reasonable, or even high, to one seller, will be such as to drive another into liquidation. In such a situation, some redistribution of capacity among the firms in the market seems to be indicated. This possibility raises broad problems of public policy on the conflict between considerations of efficiency in performance and considerations of numbers of competitors and other aspects of market structure in the application of anti-monopoly policy which cannot be pursued here.[2] Nonetheless, the assertion that any increase in cyclical price flexibility over that achieved under basing point pricing would lead to an undesirable extension of concentration requires more proof than has yet been offered.

This discussion of the relation between alternative pricing systems and the adjustment of capacity to cyclical fluctuations in demand can best be summarized by saying that there is little evidence to show that one model is preferable to another on this ground. While a pricing arrangement which led to complete price flexibility — defined as prices always equal to marginal costs at the going output rate — might have undesirable repercussions on the level of capacity in the industry, there is little to show that any of the four alternative pricing systems here examined would lead to such an extreme degree of flexibility. Aside from this, whatever advantages arise from those models which encourage more flexible pricing have already been discussed above in the section on price levels.

The last test is the effect of alternative pricing practices on the organization of industries using them. It is appropriate to examine under this head an argument made repeatedly by business men defending the basing point system: namely, that under it, market interpenetration allows larger scale production than would an f.o.b.

2. See E. S. Mason, "The Current Status of the Monopoly Problem," to be published in *Harvard Law Review*, August, 1949, for a discussion of this conflict.

mill pricing system.[3]  This argument is advanced with an earnestness and frequency that betokens firm belief in it by its proponents.  Yet it is clear that a given group of plants will have, together, the same number of customers and the same amount of demand (assuming a given level of prices) under one pricing system as under another. Differences in the allocation of these customers among the several mills, arising from the presence or absence of market interpenetration, will not affect the average output of the mills.  Thus the choice of pricing system will not in itself make either for a greater average utilization of capacity at a given scale of production or a larger average scale of production.  Nor, as was argued above, is it likely that a rigid basing point system will lead to a *lower* price level than uniform f.o.b. mill pricing — thus, perhaps, stimulating a greater expansion of demand, and allowing larger scale production.  A lowering of the price level could be expected from a system which allows market interpenetration but does not have the rigid price leadership of the basing point system (see above, pages 294–296), but this is hardly what the argument under consideration is advanced to defend.  Perhaps the business belief derives its earnestness from the individual entrepreneur's experience of the necessity of market interpenetration to his firm in periods of slack demand: he sees it as a means of getting more orders and thus increasing his production rate. Failing to appreciate the relations of his efforts in this direction to those of other entrepreneurs, he generalizes his experience to the whole industry.

The single basing point system, and the multiple basing point system with strong price leadership may well promote the growth of smaller rivals relative to the larger price leader.  These rivals are likely to be the first to shade prices if any shading is done, while the leader adheres to the formula; thus they can be more aggressive in reaching out for customers.  Against this, for the single basing point system, must be put the opportunity the price leader has for penetration into the market of any other seller without reduction in mill net. The leader may use this opportunity in such a fashion as to discourage efforts at too fast growth by smaller rivals.[4]  Under the multiple basing point system, the cost of freight absorption sets limits to the ability of the leader to penetrate other sellers' markets, and thus his

3. See for example Vol. III of *TNEC Papers* published by the United States Steel Corporation, *The Basing Point System*, especially the discussion of "local monopolies" which would arise under f.o.b. mill pricing.

4. Professor Machlup, *op. cit.*, makes much of this point.  See especially Ch. 5, pp. 151–168.  In fact, he makes somewhat too much of it, in the author's opinion.

ability to use penetration as a punitive measure. The steel industry offers an example of the decline of the price leader's share of the market which was continuous over the whole history of the firm. U. S. Steel's share of the market for steel measured in terms of ingot production declined from 66 per cent in 1901 to 33 per cent in 1938.[5] This decline went on under both Pittsburgh Plus and multiple basing point pricing, suggesting that whatever opportunities for punitive retaliation were open to the Steel Corporation under Pittsburgh Plus were not effectively utilized.

Under uniform f.o.b. mill pricing, while price leadership could continue to function, secret price cutting would not be possible, and therefore, the leader and the followers would be similarly placed in this respect. Thus large firms might maintain their relative dominance better under uniform f.o.b. than under basing point pricing. Under f.o.b. mill pricing with discrimination, a firm pattern of price leadership would not develop; a dominant producer might or might not compete aggressively against smaller rivals, and thus might or might not maintain his share of the market. This would also be true under universal freight equalization with independent base price determination.

The fact that a change in the relative shares of the market over time which lessens the dominance of the largest producer is facilitated by basing-point pricing (with price leadership) is not in itself necessarily desirable. To the extent that this change results from a kind of competition in which the level of prices and margins is little affected, whatever gains are made in terms of transferring production to more efficient units are retained by these firms, and the change produces no general benefits. If a market in which the same number of firms share more rather than less equally, is considered preferable on grounds of the undesirability of "bigness" as such, even though the change has no effect on their behavior, then the kind of change the basing point system promotes is desirable. It often happens that the firm with the largest share of the market is a long established producer with less favorable plant sites and a greater proportion of old equipment than some of his smaller rivals. Then the kind of competition promoted by the fourth model pricing system or some similar arrangement is likely not only to cause a change in the relative shares of dominant and smaller firms in the market, similar to that favored by basing point pricing, but also to result in lower costs and prices.

A multi-plant firm may find it possible to practice geographic

5. *TNEC Hearings*, Part 26, appendix, p. 13853.

price discrimination even under uniform f.o.b. mill pricing by what might be called "internal freight absorption." It could do this by shipping goods from a distant plant and billing the customer from a nearby one, or by making the actual transfer of the product from one plant to another. Another variation of this practice would be possible where output was produced in several stages; but all the final stage of production could be carried on at one location, while finishing was done elsewhere and the final product marketed from the finishing points. In either case a multi-plant producer could sell in a certain local market area with much less investment than required of a single plant producer: in the first case by operating only a small plant in the area and using excess capacity elsewhere to augment the supplies of the local plant; in the second, by operating only a finishing plant (or even a warehouse) locally, while the major part of the production took place in other areas. Such practices, by permitting larger producers to penetrate the markets of smaller ones more easily than the smaller can retaliate, might decrease the ability of local producers to survive. Under present law, however, these practices could be restrained. Either the Robinson-Patman Act, forbidding price discrimination not justified by cost differences which tend to injure competition; section 5 of the Federal Trade Commission Act, which forbids unfair methods of competition, as found by the Commission and upheld by the courts; or even, in some situations, section 2 of the Sherman Act, which makes monopolization or attempted monopolization of a market illegal, could be invoked against a large firm which used such devices to the detriment of local producers. With any pricing system which permitted market interpenetration, the local producer would be placed under no special disabilities in competing with his larger multiplant rivals. He could choose to retaliate against penetration of his market, and would have only those handicaps imposed on him by his smaller size and limited financial resources.

All in all, uniform f.o.b. mill prices appear to be no more, and possibly somewhat less, favorable to the existence and activity of independent competitors in the market than the other pricing models considered. Whatever virtues are possessed in this respect by basing point systems with price leadership are present in a stronger form in systems which allow market interpenetration but do not encourage price leadership. None of the systems offers any savings from changes in the scale of production, save as the lowering of price levels under model (4) or variants of it might stimulate a greater rate of growth of demand.

The results of all four tests taken together indicate that basing point pricing with price leadership (that has been characteristic of it) rates fairly low. In the form of a single basing point system, it receives low marks on all the tests. A universal freight equalization system with price leadership rates somewhat better in respect to the adaptation of capacity to demand in the long run, including its effects on location, and perhaps in respect to the level and flexibility of price and to the viability of independent competitors too. A change from basing point pricing to uniform f.o.b. mill pricing would not be a great improvement, judged by the foregoing tests. In respect to price flexibility and the size of margins, uniform f.o.b. mill pricing would be little better than the basing point system. As far as it affected the ability of small producers to compete successfully against large ones, it might even be a little worse. Only in two respects does uniform f.o.b. mill pricing show clear advantages over basing point pricing: it would function much more effectively in facilitating the reduction of capacity in the face of a long run decline in demand, and it would eliminate the wastes incident to the uneconomical use of transportation arising from the discouraging of road and water transport, as well as that arising from cross hauling. The situation of secularly declining demand is hardly typical for the American economy, and therefore the importance of the first advantage is not great.

On the other hand, a change from the basing point system to some pricing system which, while allowing market interpenetration, made difficult the formation of a rigid pattern of price leadership and still allowed customer choice of means of transport, could be expected to lead to fairly substantial improvements from a social point of view in the performance of the industries concerned. Most important would be the increased scope of price competition, with its effects in lowering costs and prices. Wasteful use of transportation would be avoided under this system as well as under f.o.b. mill pricing. The adaptation of capacity to changing regional patterns of demand would proceed in such a way as to permit substantial utilization of excess capacity in old regions before new capacity was added. Finally, the possibility of price as well as non-price competition would offer to new firms in favorable locations using new techniques a method of striving to displace old, well entrenched producers. This development would benefit the general public as well as the successful competitors.

What must be done to introduce a pricing system like that of model (4) in place of the now illegal basing point pricing practices? It is likely that businessmen in the industries affected would accept

it, since it would permit them to continue the practice of market interpenetration and to maintain their old customer relations, which their testimony before the Capehart Committee shows they wish to do. The effect of such a system in discouraging price leadership would not be immediately obvious, and thus would not be a reason for objection to it by the business community. The chief obstacle lies in the pronouncements of the law and of those who enforce it on price discrimination. The Robinson Patman Act, in section one, explicitly forbids price discrimination, not justified by cost and not made in good faith to meet competition, when it is likely to injure competition. The Federal Trade Commission, in its presentation of the Cement and Rigid Steel Conduit Cases, has shown itself zealous to enforce these provisions. The courts have blessed the FTC views. Further there has been a tendency in the FTC and the Anti-Trust Division of the Department of Justice to interpret "injury to competition" as injury to a competitor, as evidenced by the recent A. & P. and Standard Oil Company (of Indiana) cases.[6]

This body of law and opinion must be changed in some way if a more competitive pricing system is to replace basing point pricing. The essence of such a system involves discrimination, as was argued in section II above. For instance, a universal freight equalization system might be made legal by Congressional enactment, provided that the purchaser could, if he desires, specify the method of delivery and that he pays in freight figured into the delivered cost no more than the actual cost of delivery (though he may pay less). Such an arrangement must soon lead to occasional discrimination by sellers in favor of customers located on cheap transport routes accessible to rival sellers; and these varying discriminations will prevent the new price system from hardening in a set mold. The required change probably must be made by legislative action, since the process of changing the point of view of the courts, demanding as it does a prior conversion on the part of the anti-trust division and the FTC is always a slow one. At present, when the tide of opinion in these agencies seems to be running the other way, it would be even slower. Since there is already a demand for legislation on the subject of basing point pricing (see page 290 above), it may be hoped that vigorous discussion of the issues involved will produce an atmosphere in which constructive legislative action is both possible and likely.

6. United States v. The New York Atlantic and Pacific Tea Company, CCA 7th, February 24, 1949 and Standard Oil Company v. Federal Trade Commission, CCA 7th, March 11, 1949.

CARL KAYSEN.

SOCIETY OF FELLOWS
HARVARD UNIVERSITY

## MACHLUP ON THE BASING-POINT SYSTEM

Professor Machlup[1] has scored an academic "beat." Having under preparation a volume on the economics of price discrimination, he has pushed the chapters on basing-point pricing into publication in time to ride the crest of the wave of public interest resulting from the cement decision and the hearings of the Capehart subcommittee. He should be a powerful ally for the Federal Trade Commission forces, for he has produced the ablest book-length presentation of the most extreme anti-basing-point attitude. He follows the line consistently, plus some points that were new to this reviewer, and which bolster up shaky sectors. The reviewer is known to dissent from some of the Commission's positions, and an author has a right to a review in which criticism is tempered with a sympathetic understanding of what the author undertook to do. But the reviewer has at least attempted a sympathetic understanding of the situations and positions of both sides. And when the book is an indictment, the defendant has rights too; and so has the reading public which must try to judge the dispute. These considerations move the reviewer to continue.

Machlup's own philosophy he reveals as "old-fashioned liberalism," with faith in a "truly competitive order" and "confidence that a free society can be established and maintained by a people who insist on a wide dispersion and decentralization of economic and political power." Later, he reveals that this includes faith in a cyclically-flexible price structure — a point on which economists have more doubts than on the general credo. On this latter, there will be little dissent, though it is not clear that the ideal of competitive decentralization can do all that is required of it in an economy committed. to maintaining and stabilizing employment, with major reliance on responsible investment and pricing policies on the part of industries which presumably have some latitude of discretion in these matters, and are not completely governed by automatic and competitive forces. Our dominant standards in this matter are not unified. However, for purposes of this review, the competitive criterion will not be questioned.

Machlup discusses the matter of bias, noting that nearly all the available material is presented on behalf of opposing parties in litigation. He does not say he believes that all the bias in the litiga-

1. Fritz Machlup, *The Basing-point System: An Economic Analysis of a Controversial Pricing Practice* (Philadelphia: the Blakiston Company, 1949).

tion is on one side, but his argument proceeds entirely on that basis; and it never appears to occur to him as strange that truth should be the exclusive possession of one contending party. He does not believe in competitive decentralization of truth — in this matter at least. As to expert testimony, he notes, correctly, that it is confined to answering carefully selected questions, but does not note the basis of this in the law of evidence, whereby counsel is "bound" — whatever that may mean legally — by the testimony of his own witness, so that two-sided testimony would have a one-sided legal effect, which counsel cannot afford to risk.

The reviewer has fancied that he found distortion about equal on both sides; and that the proper function of the unattached economist starts with debunking both impartially, keeping in mind that demolishing an opponent's argument does not solve the economic problem faced by the community, and should be a preliminary to realistically-constructive analysis and recommendations. Since Machlup has done half the debunking very effectively, that would leave the reviewer the unwelcome task of doing the other half and debunking Machlup. But this would require another book. And the task is formidable; the structure Machlup is presenting is the cumulative result of twenty-five years' labor by able advocate-specialists, its semantic devices tested and improved in the arena. It often makes the testimony of company officials, temporarily diverted from the work of running their companies, look naïve and amateurish by comparison. And such officials have moments of unguarded frankness, such as their well-schooled opponents never seem to permit themselves. The latter never forget the bearing of a point on the legal strategy of their case. This advantage of continuity appears to have reached a point at which it overbalances the weight of legal talent which the companies can bring to bear. This last Machlup exaggerates, (p. 44, n. 10) with an imaginary picture of three Commission attorneys in the cement case "facing" forty-one company attorneys. The working ratio at most times was more like three to five or six, and the occasional interposition of separate counsel for particular companies was probably as much an embarrassment as an aid to the staff which had continuous charge of the case.

Without impugning the sincerity of either Machlup or those who built up the Trade Commission case, it seems impossible to understand that case without realizing how perfectly it is tailored to fit the task of outlawing a particular trade practice under statutes that do not explicitly declare that practice unlawful. The practice centers in meeting competitors' prices ("matching" those of your "co-conspir-

ators," if the reporter is on the Commission's side and employing its brand of semantics) and the statutes stress conspiracy, "unfair competition," and more recently, "discrimination in price," with a dispensation for meeting competition in good faith, which seems to have been nullified by the most recent developments. From the reviewer's standpoint, the meeting of prices is normal competitive behavior, including differential mill-nets where producers at different points sell in overlapping areas. If the meeting of prices is regular enough, it comes to be expected, and so leads to the effect analyzed in oligopoly theory, with its characteristic sticky prices and mitigated competitive pressures. Those in the business recognize this and promote regularity, by upholding it as fair practice and condemning departures, by joint implementing action (e.g., common freight-rate books, converted into the units in which prices are quoted) or by the action of one producer (as when an irregular situation is met by announcing regular prices as low as, or lower than, the irregular ones previously in effect). The Commission, wanting to outlaw any "systematic" meeting of prices, finds these supporting practices helpful as evidence of collusion or coercion, but has pushed on toward the culminating doctrine of collusion inferred from mere conscious similarity of conduct. All this is necessary because, while the main thing objected to is the stalemating effect of knowing that a price reduction will be instantly met, this effect is not, *per se*, illegal. It has to be attacked as an end sought through combined action or discrimination.

All this concern with whether past practices violated the language of a statute is necessary, but often seems to get far from the question of the economic effect of the practice for the present and the future, especially as judged by whether the effect of outlawing it is likely to bring any improvement. And that is one final reason for not undertaking all the debunking which one might be tempted to attempt.

Machlup's basic picture is one of industries dominated by a few great concerns, using basing-point pricing as one among various deliberate devices for extinguishing competition in price (competition in service is neglected, and competition in quality is perhaps vestigial in these standard-product industries). This conscious control spells oligopoly in a broader sense than the sticky-price effect already mentioned, which can take place without collusion or conscious controlling action. This control is held responsible for the growth of industry taking place mainly through increased size of firms and only in small part through increased numbers (p. 167). Basing-point pricing is represented as playing its part in this discouragement of

new entries and retardation of their growth, in contrast to the more usual picture of it as part of a policy that "holds an umbrella" over the smaller competitors and allows them to grow faster than the company that bears the brunt of stabilizing prices. Machlup does not even mention the "umbrella" theory, nor the figures of relative growth which have given support to it.

His picture is one of the big concerns exerting coercive pressure on smaller outlying concerns to remain non-base mills, or to keep their base prices high, thus restricting their growth; when it would be to their individual interest to set lower prices (pp. 159, 166–7, and cf. p. 101, l. 9). "A small price reduction could thus secure him a most remarkable increase in volume," "provided his competitors did not retaliate" (p. 112). "Retaliate" is semantics for any attempt to hold their customers and trade area when a competitor starts to take them away. But if they did thus fail to defend their competitive interests, one wonders how long it would be before it would be evident to Machlup and the Commission attorneys that their conduct was non-competitive. At other points (pp. 103, 176–7) Machlup admits the possibility of non-collusive freight absorption, which "may in certain situations (situations not specified), promote price competition" but thinks that, through retaliation or fear of retaliation, it would tend to eventuate in some collusive arrangement for the restraint of the practice. He does not mention that the practice might be restrained because sellers naturally don't like to absorb more freight than they have to, to secure the volume they think they need. Here, as elsewhere, he serves his argument by blurring the distinction between punitive reprisal and normal competitive effort. At another point (pp. 118–20) he sees nothing "natural" about basing-point pricing except that it is natural to want to get rid of competition.

At still another point (p. 169) he says: "Without collusion the competing producers would try to gain bigger shares of the high-price business and, in competing for it, would bid down the higher prices until they were no higher than any others." This might have some force as an argument that non-base mills would not remain non-base mills; but as an argument against the naturalness of freight-absorption by base mills, it breaks down. The actual situation of a competitor with no monopolistic intentions is something Machlup nowhere contemplates long enough to ask what such a one would "naturally" do, for example, if he were setting up the first cement plant in the Mississippi basin, and deciding how to meet the sales from the well-established Lehigh Valley district; and what he would expect the

Lehigh producers to do, in the normal effort to hold their existing customers. There is no recognition of the consequences of the fact that a general price reduction, while it may bring added sales, reduces the revenue from the existing volume of sales, while the immediate gain in revenue from an additional freight-absorbing sale is subject to no such deduction. Accordingly, his picture of what would happen in the absence of monopolistic collusion leaves out the principal reasons for freight-absorption. He avoids the explicit claim that, but for collusion, uniform f.o.b. mill pricing would naturally establish itself in these industries; but he gets the benefit of arguments that rest on this unwarranted claim (cf. pp. 102–6, where he takes this as his standard of comparison by which to judge the effects of basing-point pricing).

In his picture, freight absorption (usually objected to on the ground that it is not strongly enough competitive) shifts at will to the other extreme and becomes a "built-in" instrument for coercive reprisal, the modern substitute for the localized predatory raiding of early Standard Oil days. The salesmen of the larger concern are represented as being able at will, with no change of prices, to multiply their sales in a rival's home area, merely on a hint from the home office that it would look favorably on such expansion. Machlup does not say that freight absorption has actually been used in this way, merely that it could be so used; but that seems sufficient to convict it in his mind.

This argument probably underlies the identification of the basing-point *form of price structure* with agreement to fix the *level of prices*, intimidating any who would reduce them. The origins of the practice, as well as its ultimate form, are construed as the devising of a "formula" — presented with unnecessary wonder at its alleged ingenuity — for the deliberate purpose of extinguishing price competition. As to origins, the industries would vigorously dispute the picture offered.

Machlup seems to think that the companies knew from the start that they were violating the law, though elsewhere he says a thing is not illegal until it has been so declared (that is, not till 1945–48) and puts the blame on the authorities for slowness in enforcing (declaring?) the law. The law must be flexible — here the reviewer agrees — and: "To complain about the 'uncertainty' which this involves is really to complain about the risks of evading the law" (p. 37). Here in one short sentence he both asserts and denies the uncertainty of the law. The logic of this really amazing sentence is to assert that all doubts are settled in advance against the companies.

Only when ready to discuss the effects of outlawing the system does Machlup explicitly assert that the effects of a practice can be judged only by comparison with possible alternatives.  He mentions several possible remedies which he dismisses as "untried and rather vague proposals" (p. 184).  They are, be it noted, no vaguer, and are less ambiguous, than the concepts of what competition would bring about in the absence of basing-points or of collusion.  Yet these are required by his logic as a basis for any statements about the effects of basing-point pricing, and as to these he has no doubts.  And he speaks as if the outlawing of basing-point pricing had definite meaning, though it sets up no defined alternative, and therefore, by his stand-ard, is meaningless.  For this purpose he concentrates on uniform f.o.b. mill pricing, not claiming that this is what the present law requires, but holding that nothing short of this remedy is sufficient.  He refutes the claim that it would set up "local monopolies," ignoring its effect on competition in service, with respect to which there seems to be something in the "local monopoly" idea.  And he dismisses the claim that uniform f.o.b. mill selling would not affect the degree of price competition, simply ignoring the rather substantial arguments that have been made for this contention (p. 185).

His standard of desirable price behavior, which he expects this system to further, is flexibility, with prices lower in depressions and higher in booms than they have usually been.  But apparently the boom of 1948 is an exception; he sees no reason why the introduction of the system in that year should not have led to lower average prices.  Instead, base prices were not changed; and former freight-absorbing business now paid full freight, so these delivered prices rose.  Machlup objects that in this the steel industry acted from political rather than economic motives, attempting to discredit the new system.  He holds that it would lead to larger output and less unused capacity.  This is a two-sided question.  To the extent that fluctuations of demand in each region have to be supplied by the producers in that region, the average fluctuation of producer-output would be increased, and more reserve capacity be required.  But if (as is doubtful) prices were driven lower in dull times, so that profits could be earned only when demand was close to capacity, the incentive to limit capacity would be increased.

The major difficulty is that the effects of uniform f.o.b. mill pricing are highly uncertain.  Some can be tentatively anticipated by serious study, but this is not a task to be quickly and easily accom-plished, either by seeking advice from theoretical economists or by inviting interested parties to testify before a Senate subcommittee,

although both methods may be useful in suggesting angles for further investigation. By comparison, Machlup's conclusions are seen to rest on an inadequate base; there are important and easily-foreseeable complications, mentioned in the literature of the subject, with which his argument does not reckon. The considerations involved in the location of basic industry, fabrication, and trade, and in shifts of location, would repay intensive study; also the effects of stable-employment policies on this problem. Worth investigating is the hypothesis that less drastic and expensive changes offer more prospect of strengthening price competition than does compulsory uniform mill pricing, and might be more effective without the latter than with it.

The reviewer, while consciously trying to offset Machlup's bias, has endeavored not to be unfair, and would be genuinely glad to be shown points at which he has been less than just. Like Machlup, he takes the problem seriously; so much so that he would rather acquiesce in some rulings from which he dissents, than see the power of control emasculated, if that were the alternative. But if oligopoly is as prevalent and as firmly rooted as Machlup represents it, dislodging or substantially weakening it will take more than even a drastic change in geographical pricing practices. Machlup's picture of conditions would prove his own remedy inadequate.

J. M. CLARK.

COLUMBIA UNIVERSITY

# THE POST-WAR REORGANIZATION
## OF THE GERMAN BANKING SYSTEM[1]

SUMMARY

I. Basic purposes, 322. — II. Attempts at quadripartite agreement, 325. — III. Reorganization in the U. S. Zone, 325. — IV. Reorganization in the French and British Zones, 328. — V. The new Central Bank in the Western Zones, 330. — VI. Reorganization in the Soviet Zone, 333. — VII. The situation in Berlin, 337. — VIII. Conclusion, 339.

## I

### BASIC PURPOSES

The general principles for the reorganization of the German financial system were laid down in the Potsdam Agreement of August, 1945. The Agreement provided, on the one hand, that "the German economy shall be decentralized for the purpose of eliminating the present excessive concentration of economic power"; and, on the other, that "Germany shall be treated as a single economic unit" and that, "to this end common policies shall be established in regard to . . . currency and banking." This continued to be the basic policy, for as late as July, 1947, the Commander-in-Chief of the American Forces of Occupation in Germany was directed by his government to "take such action as may be necessary to prevent the establishment of a centralized German banking system and an undue concentration of financial power, but to encourage the establishment of a central authority for the production, issuance and control of currency and for technical banking supervision."

The application of these principles to Germany's banking system, as interpreted by U. S. Military Government, involved five basic policies[2]:

1. Decentralization to a Land[3] basis of banking and of the routine supervision of banks;

2. Establishment of a central bank for each Land;

1. The author was, during the major part of the period discussed, chief of the Banking Section of the Office of Military Government for Germany (U. S.). The views expressed are, of course, his personal ones, and in no way reflect the views of that organization.

2. These policies were laid down in late 1945 by Joseph M. Dodge, later president of the American Bankers' Association, then director of the Finance Division, Office of Military Government for Germany (U. S.).

3. A German Land is similar to a State. Postwar Germany is divided into sixteen Laender.

3. Establishment of an Allied Banking Board and a Land Central Bank Commission for the purpose of controlling and administering the German currency and banking system by coordinating the activities of the Land Central Banks;

4. Elimination of the Reichsbank and the central offices of the "Grossbanken"[4];

5. Elimination of excessive control by the banking system over German industry.

The decentralization of banking was to be carried out by prohibiting any bank from establishing or retaining branches outside the Land in which its head office was located.[5] Banks were to be given a choice as to the Land in which they preferred to establish their domicile; branches and offices in all other Laender were to be liquidated or transferred to independent purchasers. Banks were to be completely independent of any control by other banks, government agencies, bank associations, central associations,[6] or any similar organizations located outside the Land in which the bank was located. However, this reorganization of Germany's banking system was limited to the decentralization of its control and organization, and was not to preclude any bank from carrying on business outside the Land in which it was located, nor from establishing correspondent bank relations with banks located outside the Land. The general supervision over banks was to be a function of the finance authorities in each Land.

A central bank was to be established in each Land. Its capital was to be provided initially by the Land government, but was to be subsequently reapportioned among the individual banks of the Land. The Land Central Banks were to serve as the channel for the distribution of Germany's currency, provide credit facilities to the individual banks in the Land, regulate interest and discount rates, execute open-market operations, regulate the reserve requirements of banks and hold such reserves in the form of deposits, serve as fiscal agent for the Land government, and supervise the clearing and settlement of financial transactions of the banks within the Land.

The activities of the Land Central Banks were to be coordinated by a Land Central Bank Commission, representing the central banks of the Laender, and an Allied Banking Board, representing the Allied

---

4. This term refers to the six biggest German banks.

5. Original plans called for even further decentralization to county and township level, in accordance with the so-called "Morgenthau-Bernstein" policy.

6. For example, Sparkassen- and Giroverbaende, Girozentralen, and Zentralkassen.

Governments.  The Allied Banking Board and the Land Central Bank Commission were to establish common policies with respect to currency and banking in order to achieve maximum uniformity in the individual Laender, and were to be the sole authority in Germany for the issuance of currency.  In their relation with the Land Central Banks, the Board and the Commission were to have sole control over the distribution of currency, the general regulation of bank credit, including interest and discount rates, the investment and open-market policies of the Land Central Banks, and the regulation of reserve requirements of the Land Central Banks and all other banks. Furthermore, the Board and the Commission were to control the institutions responsible for the clearing and settlement of financial transactions above Land level, and regulate all foreign exchange transactions.

The Reichsbank and its affiliates were to be liquidated.  The head offices of the "Big Six" — Deutsche Bank, Dresdner Bank, Commerzbank, Bank der Deutschen Arbeit, Reichskreditgesellschaft, and Berliner Handelsgesellschaft — were also to be liquidated, and their branches were to be reorganized on an independent basis within each Land.  The Bank der Deutschen Arbeit, as the bank of the German Labor Front, was to be dissolved.  The number of banks and branches was to be severely cut in order to adjust the banking apparatus to the drastically reduced requirements of the German economy.

The elimination of excessive control exercised by the banking system over German industry was to be achieved by severing some of the interrelationships between the banks and large corporations. The control had been facilitated through investments in the stocks of corporations, as well as by the fact that in Germany banks acted as principal agents in handling stock exchange transactions, floating new securities, and sponsoring mergers and other types of combinations.  Banks voted not only stocks owned by them, but also those held by them for safekeeping only.  It was planned to forbid banks to be members of any stock exchange or to buy or sell stocks or shares except as agents for their customers, and then only through brokers or other established agencies other than banks.  Banks were not to be permitted to invest their funds in common or preferred stock of corporations, and under no circumstances were they to exercise any voting rights in respect of any stocks or shares of any commercial or industrial company.  There were, however, no definite plans to eliminate other methods of control, such as, for example, interlocking directorates.

## II. Attempts at Quadripartite Agreement

This American plan was introduced in November, 1945 for discussion at the Allied Control Authority, the quadripartite body governing Germany. In the early discussions, the main objections came from the British delegate. According to the British view, a radical reform of the banking system was not necessary, since it was an efficient system and did not wield excessive power; the decentralization would make Allied control more difficult, make the economy more sensitive to crises and failures, and facilitate the tendency toward "local monopolies." On the other hand, the Soviet and, at first, the French delegate, too, felt that decentralization was the most urgent problem and that the creation of a central banking mechanism should be postponed until the decentralization had been carried out.

In the course of the discussions, the various views shifted considerably. When the matter came up for a final discussion at the Allied Control Council in October, 1946, the U. S., British and French Military Governors agreed to the American plan for the reorganization of the German banking system, though the British Military Governor made it clear that he could accept the decentralization provisions only if the "constructive" part of the plan — the establishment of a central banking mechanism — was simultaneously agreed to. The Russian delegate accused the western powers of using the proposed legislation for preventing the decentralization of "bank monopolies." He was willing to agree to the decentralization provisions of the plan, but not to the establishment of a central organ. This ended the discussions at the Allied Control Authority, and each delegate advised his Government that no agreement could be reached.[7]

### III. Reorganization in the U. S. Zone

Simultaneous with the first quadripartite discussions on the subject late in 1945, the U. S. Military Government informed the German Laender governments in the U. S. Zone of the basic principles that were to govern the reorganization of the German banking system. It was hoped that the German authorities could be convinced of the desirability of such a reorganization and put it into effect as their own legislation. Not only did this hope turn out to be vain, but in the meantime, the German banks were making strenuous efforts to

7. Legislation for the elimination of excessive control exercised by the banking system over German industry had already been introduced by the U. S. delegate at the Allied Control Authority in October 1945. The matter was discussed innumerable times, but no agreement had been reached when the Allied Control Authority ceased effective operation.

reestablish their previous positions and reintroduce the centralist controls torn apart in the last months of the war. The head offices of the Deutsche Bank and Dresdner Bank, which had been closed in Berlin, were replaced by central offices at Hamburg in the British Zone; the Commerzbank already had a head office in Hamburg. They now attempted to control from there the remnants of their systems in the western zones in the same manner as they formerly did from Berlin. As a result, U. S. Military Government felt itself forced, in March 1946, to instruct banks located in its zone that they must neither ask for nor accept instructions from any banking office outside the U. S. Zone.[8] This order was considered to be of a temporary nature only, pending a more permanent law to be written, it was still hoped, by the German authorities themselves.

However, more than a year later, in May, 1947, when the German governments had failed to take, or even to plan, any positive steps, when quadripartite action seemed unlikely, and when the banks were still making strenuous efforts to reassert centralized control, the U. S. Military Government was forced into a position of ordering the decentralization of the control of the Deutsche Bank, Dresdner Bank and Commerzbank.[9] Independent and disinterested custodians were appointed for these three "Grossbanken" in each Land. The custodians, responsible to the Land governments, were to manage and administer the property of these banks within each Land; but, in so doing, they were not to be influenced by the shareholders or officials of these banks. The names of the banks were changed so that the new names differed in each Land; twelve new names, three for each Land, replaced the previous three names.[1] Their giro-systems[2] were broken up by requiring large clearings above Land level to go through the newly established Central Banks of each Land. The existing relations between these banks and their customers were not altered and their financial position in each Land remained unchanged. It was felt that a final reorganization of the ownership — in contrast to control — of these banks into separate, independent

8. Even though the order referred to a decentralization of control to a zonal level only, it was interpreted widely as decentralization to a Land level.

9. *Military Government Law No. 57 — Appointment of Custodians for Certain Bank Organizations.* No action was taken in regard to the Reichskreditgesellschaft and the Berliner Handelsgesellschaft, which maintain no branches. The liquidation of the Bank der Deutschen Arbeit had already been agreed to by the Allied Control Authority and its closing was carried out in all four zones.

1. It was expected that currency reform would force considerable consolidation among these banks, as well as among the banking system as a whole.

2. In Germany, the giro-system is a bank transfer system for effecting checkless transfers of funds through a clearing system.

regional institutions should await the outcome of quadripartite discussions on the liquidation of their Berlin head offices, and that, in any case, the necessary evaluation of the property would be difficult before a financial reform.[3] Even though German officials had warned that the decentralization would weaken public confidence in the banking system and might lead to its collapse, there was in fact no run on the banks, no increase in withdrawals, no attempt to transfer funds into other zones, and the stock quotations of the affected banks, after a temporary decline, reached new heights a few months afterwards.

The lack of cooperation on the part of the German authorities in the decentralization of the "Grossbanken" was matched by their failure to submit adequate plans for the replacement of the Reichsbank by central banks in each Land. Proposals submitted to Military Government dealt mainly with the undesirability of undertaking the reorganization and requested its postponement. Beyond this, the main difference between the American and German authorities was on the degree of political control to be exercised by the governments. It was the German contention that the banks should be owned and controlled by the state, in contrast to the American policy of having the banks owned by the member banks and controlled by representatives of the economy as a whole, including the government. The American position prevailed and appropriate legislation was promulgated in December, 1946, stating specifically that it was requested by Military Government.

A central bank was established in each Land of the U. S. Zone of Occupation on January 1, 1947.[4] While the ownership of each Land Central Bank was, at first, in the hands of the Land governments, the stock was to be sold within two years to the various banks operating within the Land.[5] Determination of major policies was in the hands of a board of directors, consisting of representatives of trade and industry, agriculture, labor, banking and the government. The banks took over the main functions of the former Reichsbank, though their clientele was limited to their member banks; for each central bank was conceived as a bank for banks and not as a competitor

3. About three-fourths of the assets of these banks consisted of government securities.

4. They were located in Munich (Bavaria), Stuttgart (Wuerttemberg-Baden) and Wiesbaden (Hesse). The Land Central Bank of Bremen was established three months later, by which time Bremen had become a Land of the U. S. Zone.

5. This period was extended late in 1948, so that the stock is still in the hands of the governments.

of commercial banks.   Among their major functions were insuring the
solvency of credit institutions, keeping available reserves against
deposits, engaging in open-market operations, execution of cash
transactions for the Land Government and granting of short-term
credits to public institutions where this was not performed by other
institutions, attending to non-cash transfers and the clearing of
checks, and assisting in financial transactions with other German
Laender.   The Land Central Banks were to regulate the circulation
of currency and the supply of credit in each Land; but they were not
given the right to issue currency, though it was contemplated that
they were to become, in time, the channel through which currency
was to be distributed.[6]

The activities of the four Land Central Banks were coordinated
until early 1948 by an Advisory Council consisting of two repre-
sentatives of each central bank and one of each Land Ministry of
Finance.   The functions of the Advisory Council were limited to the
establishment of common policies in the field of interest and discount
rates, minimum reserve requirements, open-market policies, and
transfer and check transactions.   Its recommendations were, however,
not binding on each Land Central Bank.[7]   It was hoped that an
expansion of the Advisory Council would be able to serve, in time,
as the central banking agency for all of Germany.

## IV.  REORGANIZATION IN THE FRENCH AND BRITISH ZONES

The reorganization of the banking system in the U. S. Zone was
followed by the nearly parallel action in French Zone.   German
banks were permitted to continue functioning and, with the exception
of the branches of the Bank der Deutschen Arbeit, no banks were
closed.   Banks whose head offices were located outside the French
Zone were requested to sever connections with their head offices and
Military Government representatives were appointed to supervise
the activities of the branches of each of the banks concerned.   In

6. While the new banks were not the legal successor of the Reichsbank,
they took over the current assets of the Reichsbank within each Land and assumed
the liabilities of the Reichsbank as far as it was possible.   Since claims on Berlin
originating out of pre-occupation business were not transferred to the Land
Central Banks, the resultant excess of liabilities over assets was balanced by a
claim on the trustee of the Reichsbank for the Land.   A general trustee for the
Reichsbank was appointed for the entire U. S. Zone in order to facilitate the
eventual liquidation of the Reichsbank.

7. This was due to an error in translation.   In accordance with the intention
of Military Government, the recommendations of the Advisory Council were
binding in the English text, which was, however, not the legal text.

March, 1947, Land Central Banks were established in each Land of the French Zone, which were nearly identical to the ones established three months earlier in the U. S. Zone. The only significant differences were that the question of ownership of the central banks was left to the discretion of each Land, and that they were permitted to make short-term loans to private organizations operating in the public interest; the activities were not so strictly limited to those of a "bank for banks" as in the U. S. Zone. In October, 1947, the control of the Deutsche Bank, Dresdner Bank and Commerzbank was decentralized in the French Zone in the same manner as had been done in the U. S. Zone the previous May. The names of these banks were changed, their giro-system cut, and custodians appointed to eliminate the control of the central offices in Hamburg. No changes were made in the ownership of the banks.

In the British Zone, unlike the other three zones, the banking system remained substantially unchanged until 1948. In the British view, decentralization of banking without some form of centralized control was not a workable solution and not in accordance with the principle of economic unity established in the Potsdam Agreement. Centralized supervision over the banking system was carried out directly by the British Military Government itself. Not until January, 1947 was some supervision — but excluding supervision over central banking, currency, reserve requirements and foreign exchange — put into the hands of the Land governments.

Zonal central offices were established in Hamburg for the Reichsbank, Deutsche Bank, Dresdner Bank and Commerzbank to replace, at least in part, the former head offices in Berlin. Though these central offices were not supposed to exercise any control over their branches outside the British Zone, they attempted constantly to reassert their previous position. When, for example, instructions from these offices were no longer permitted as orders, they were quite effectively forwarded as "information."

In spite of British opposition to the decentralization, the action taken in the other three zones automatically decentralized the banking system of the British Zone to a zonal level. Beyond differences of principle, the main difference between the situation in the U. S. and British Zones was that, while the decentralization in the U. S. Zone had been carried to a Land level, that in the British Zone was carried only to a zonal level. The economic fusion of the two zones, as well as the *de facto* situation created in the British Zone as a result of the action taken in the other zones, made an agreement between the two Military Governments possible in October, 1947. In this agreement,

the British accepted the decentralization of the "Grossbanken" and the replacement of the Reichsbank by a Land Central Bank system, · both of which were finally carried out by the spring of 1948. In return, it was agreed to establish a central agency to coordinate the activities of the eight Land Central Banks of the two zones.

## V. THE NEW CENTRAL BANK IN THE WESTERN ZONES

During the discussions at the Allied Control Authority, it had been the American position that Germany's decentralized banking structure be coordinated by a board consisting of representatives of the Land Central Banks. However, in the course of the bipartite discussions with British Military Government in the fall of 1947, this position was basically altered. The new central agency was not to be a board, but a central bank.[8]

On October 30, 1947, agreement was reached that a central bank would be established in the West. In the agreement, the board of directors is composed of the presidents of the Land Central Banks, and a chairman and president of the board of managers selected by them. The directors exercise sole control with respect to the basic policies and functions of the Bank. It establishes common policies with respect to currency and banking and maximum possible uniformity of banking policies of the individual Laender. It issues and distributes currency,[9] provides for the general regulation of bank credit, including interest and discount rates and the investments and open-market operations of the Land Central Banks, rediscounts paper for them, regulates reserve requirements, holds deposits (including reserves), regulates the custody of securities, assumes the clearing and settlement of financial transactions beyond Land boundaries, serves as fiscal agent for the Bizonal Economic Administration, and supervises the agency controlling foreign exchange.[1]

After lengthy discussions between the British, French and U. S. Military Governments, as well as German authorities, the "Bank

8. This change in the American position was the desire of the Military Governor, General Lucius D. Clay, though his financial advisers opposed it. The directive of the Joint Chiefs of Staff of July 11, 1947, approved by the State, War and Navy Departments, provided against "the establishment of a centralized German banking system," but only for "the establishment of a central authority for the production, issuance and control of currency and for technical banking supervision."

9. The issuance of currency is to be only on special instruction of the appropriate Allied authorities. On June 20, 1948, concurrently with currency reform, the Bank was authorized to issue Germany's new currency.

1. This supervision is to be exercised as soon as Military Government relinquishes its control over foreign exchange.

Deutscher Laender" was established, on March 1, 1948[2] in Frankfurt, as the central bank for the western zones of Germany.[3] The organization and functions of the Bank followed closely the agreed U. S.-U. K. principles of the previous October. The Bank is the "bank of the central banks" and not a competitor of existing financial institutions. It is not allowed to maintain branches or subsidiaries. Its activities are limited to transactions with member Land Central Banks, central banks of other German Laender and of foreign countries, and the Bizonal Economic Administration. Its capital of RM 100 million was subscribed by the member Land Central Banks in proportion to their deposits; annual profits, after the establishment of certain reserves, are distributed to each Land Central Bank in proportion to its capital participation. An Allied Bank Commission provides the necessary policy supervision on behalf of the Military Governments.[4]

By 1948, quite a few of the reorganization ideas originally planned in 1945 had been dropped or modified. Among the four main differences which still remain between the old Reichsbank system and the Bank Deutscher Laender is the latter's decentralized control. The decentralized composition of its board of directors assures that the policies of the Bank are directed, in accordance with the principles of a federal organization, by representatives selected by the governments of the Laender. A second important difference is that the Bank, unlike the Reichsbank, is not permitted to engage in commercial business. The introduction of minimum reserve requirements is also a new feature in the German banking system, which was not considered desirable by the German authorities consulted in the establishment of the Bank. The Reichsbank had powers to regulate bank credit, including interest and discount rates and open-market operations, but had made no use of its power to establish minimum reserve requirements for individual banks. The fourth important innovation is the assumption by the Bank of the entire settlement of bank balances resulting from transfers between the Laender. The German banking system had contained at least four different transfer and settlement systems.[5] A reduction had

2. The French Zone joined one month later.
3. *United States Military Government Law No. 60* and *British Military Government Ordinance No. 129.*
4. The creation of this Commission was necessitated in part by the insistence of French authorities, who were willing to join the organization if definite Military Government supervision were provided.
5. The four most important giro-systems were those of the Reichsbank, Reichspost, savings-banks, and cooperatives.

been recommended in the thirties, but was never carried out. They were now to be consolidated into one, insofar as transactions above Land level were concerned.[6]

Many Germans saw in the establishment of a central bank for the western zones a further step splitting Germany into two separate political entities. As such, they welcomed it no more than they did the Bizonal Economic Administration, in which they had seen a similar danger.[7] Their main technical objections turned upon two points. Firstly, they objected to the decentralized control of the Bank and the absence of control by the Bizonal Economic Administration. They felt that, while the interests of the Laender are sufficiently represented, the interests of the central government, as well as the interests of the various sectors of the economy, are not adequately represented. A second objection was the exclusion of the Bank from commercial activities. It was argued that the Reichsbank was permitted to undertake such activities and that even the Federal Reserve Banks had received permission in the early thirties to make certain commercial loans. However, these objections were not accepted by Military Government. Their acceptance would have resulted in a central banking mechanism very similar to the former Reichsbank system, and, thereby, would have practically eliminated the justification for the reorganization as a whole.

The Soviet authorities were quick to point out to the German public that the western powers had taken a further step in the division of Germany. Supposed justification for this claim was found in the right of the Bank to issue currency, which would facilitate a separate currency reform in the western zones. It was argued that the Bank Deutscher Laender was based on most conservative principles of private enterprise — contrary to the world-wide trend toward greater influence of the state in financial affairs — in order to leave the Bank in the hands of "private monopoly capital" and exclude "the people" from the supervision of the Bank. The Bank was considered a "tool of Wall Street," which would be able to control the German economy through it. The Potsdam Agreement was, of course, considered broken again.[8]

6. It seems quite possible that this consolidation will not be carried out as planned.

7. The Social Democratic Party even refused to take an active part in bizonal organizations.

8. The Soviet views were expressed in the *Taegliche Rundschau,* official Berlin newspaper of the Soviet Military Administration, February 21, 1948. I. Zvezdin, writing in the *New Times,* February 11, 1948, concludes that "using the new bank as an agent, or when necessary, as a screen or dummy, American monopoly capital will have every opportunity to seize key positions in Bizonia."

## VI. REORGANIZATION IN THE SOVIET ZONE

The reorganization of banking was most drastic in the Soviet Zone. With the beginning of the occupation in May, 1945, all banks were immediately closed and all accounts blocked. Outpayments of RM 300 were permitted to those depositors whose total account was less than RM 3,000; depositors with more than that amount were considered undeserving capitalists and received no funds.[9] The private banking system was replaced by five big state banks, one for each of the provinces[1] of the Soviet Zone, and by new city banks. Municipal savings banks and agricultural credit cooperatives were permitted to continue functioning to the extent that they acquired new deposits. Eleven small private banks were temporarily permitted to do new business.[2] A special bank, the Garantie- und Kreditbank, was established to finance undertakings of particular interest to the Soviet Military Administration.

Steps were immediately taken to "liquidate" the old banks in the Soviet Zone. In contrast to western concepts of liquidation, where available assets are distributed among the various claimants according to a definite priority, the Soviet idea of liquidation involves simply the locking of the doors of the institution after books, documents, securities, cash, etc. have been removed, and the taking over of existing assets by the new state bank, no consideration being given to liabilities. Accordingly, the old banks in the Soviet Zone were completely emptied and their available assets transferred without compensation to the five state banks, which were charged with directing their "liquidation."

The new state banks at once became the most important banks

9. The partial unblocking was ordered by the Soviet Military Administration on March 9, 1946. The order begins as follows: "Owing to the bankruptcy caused by the criminal policy of the Hitler regime, the old German banks, private and public, have no resources that would enable them to pay out any old deposits. Considering, however, the fact that the interest of the owners of small accounts have seriously suffered as a result of the bankruptcy of private and public banks, and since I deem it necessary to give them material assistance, I order . . . ." An outpayment of RM 400 was permitted to persons unable to work and lacking any other means for their support, even if their account exceeded RM 3,000.

1. The term "Province" is identical to the term "Land" in the western zones.

2. The number of private banks seems to have fluctuated during the period. The Soviet member of the Finance Directorate, Allied Control Authority, reported in July 1946 that there were nine private banks in the Soviet Zone; in January 1947 he reported the existence of eleven. Their average assets on March 31, 1946 were about RM 2.6 million, in contrast to the average assets of the state banks of RM 1.8 billion. Early in 1948, plans for the discontinuance of private banks in the Soviet Zone were reported in the German press; *Tagesspiegel*, February 25, 1948.

of the Zone, accounting by the end of 1946 for about seventy-five per cent of total bank assets.   They are completely owned and controlled by the state, the government of the province providing the capital, and the Minister President being responsible for the appointment of both the board of directors and the board of managers. The board of directors consists of the President of the province, as well as the Finance Minister and one representative each of agriculture, trade, commerce, the Ministries of Finance, Economics, Trade, Municipal Administration, Agriculture and Labor, and two members selected by the employees of the Bank.   The banks are ordered not to operate on the basis of profit and loss calculations, but in the "public interest," as defined by a central plan.   They are "Universalbanken" in that they are authorized to undertake any and all banking business, including at first even the distribution of currency. The profits of the banks, after provision for certain reserves, go to the state; any losses not covered by reserves and in excess of ten per cent of the capital are made up from the state budget.   It is interesting that even though the *raison d'être* for capital and reserves no longer exists in these organizations, it must have been hoped that organizing them along these lines would give them a semblance of business-like institutions.

The Garantie- und Kreditbank was established as the main instrument of the Soviet Military Administration for the financial penetration of the Eastern Zone.   The bank was originally founded in 1923, but after the end of the war its ownership was transferred to Soviet governmental agencies.   Even though it continued to be organized and registered according to German law and as a German bank, both its board of directors and its board of managers consisted of Soviet citizens.   In a very short time, it became one of the most important banks of the Soviet Zone and Berlin.   Its assets increased from RM 900 million in December 1946 to more than RM 2 billion by June 1947.   The bank finances the business of the Soviet-owned corporations of the Soviet Zone, which control an important part of the industrial production of the area; it has exclusive rights to handle all funds invested by the Soviet Military Administration in factories and other enterprises; it is the only export-import bank in the Soviet Zone and finances the exchange of goods between the eastern and western zones of Germany; it grants intermediate credits for the production of goods to be delivered to the Soviet Union as reparations; it has large holdings of German real estate in Berlin and the Soviet Zone; and it seems to be the only commercial bank in the Soviet Zone with branches in all its provinces.

In May 1947, "Emissions- und Girobanken" were established in the Soviet Zone with functions nearly identical to those of the Land Central Banks of the western zones, including the function of note issue. The most important differences between the two systems result from the fact that the economies served by them are completely divergent: The Soviet Zone represents a socialized, totalitarian economy, while the western economies represent a state-controlled form of capitalism, with a free enterprise system as its goal as soon as more normal economic conditions are reestablished. As a result, the significance of monetary and fiscal policies is quite different in the two areas.

Beyond this, the Emissions- und Girobanken differ from the Land Central Banks on two main points: ownership and control. The central banks of the Soviet Zone are state institutions, the provincial government providing the capital. The governments are liable for the banks' commitments, receive their net profits, and bear their losses. Even Soviet acceptance of the terms of the western Land Central Bank legislation would not have altered actual owner- ship of the central banks. For provision of the capital by the banks of the province, instead of by the government, would still have left the central banks indirectly state-owned since the state banks in each province would contribute the vast majority of the capital. The capital of the Emissions- und Girobanken would, in that case, be indirectly owned by the governments through their ownership of the state banks.

More important than the differences in ownership are differences in control. While the Land Central Bank system of the western zones recognizes the interest of the state in the banking system, the control of the banks is largely in private hands, definite care being taken that governmental control does not become dominant. The reorganization in the Soviet Zone has put the Emissions- und Giro- banken under the complete control of the state. The board of man- agers is appointed by the Minister President of the province, as are six of the nine members of the board of directors. Furthermore, not only are the central banks controlled by the governments of the respective provinces, but, until May 1948, the Central German Administration of Finance for the Soviet Zone controlled interest rates, reserve requirements, open-market policies, and carried out all other functions which in the Land Central Bank system were originally carried out by the Advisory Council and later by the Bank Deutscher Laender. Thus, while the central banking mechanism in

the western zones is organized from the bottom up, that of the Soviet Zone is controlled from the top down.

This basic situation was not changed when in May 1948, about three months after the establishment of the Bank Deutscher Laender as the central bank for Western Germany, the "Deutsche Emissions-und Girobank" was established in Potsdam as the central bank for the Soviet Zone.[3] It has the usual functions of a central bank, which it carries out by directing and coordinating the activities of the five Emissions- und Girobanken. The Bank controls the circulation of currency as well as the transfer of funds; it effects all interzonal clearings and directs all foreign exchange transactions. While the amount of its capital is identical to that of the Bank Deutscher Laender, more than one-half of it is contributed by the Central German Administration for the Soviet Zone, the remainder by the government-owned Emissions- und Girobanken. Control over the Bank remains in the hands of the Central German Administration of Finance, whose chief serves as chairman of the board of directors. Among other members of the board of directors are the presidents of the Emissions- und Girobanken and a representative of the monopoly labor union (Freier Deutscher Gewerkschaftsbund).

In spite of the great differences between the central banking system in the East and West, the purely technical problems of their coordination are not insurmountable. The question of ownership could be left to the choice of each Land, as it is, in fact, in the Land Central Bank legislation of the French Zone. If the controls of the German Administration of Finance for the Soviet Zone and the Deutsche Notenbank were transferred to an institution combining the central banks of Eastern and Western Germany, the technical differences between the eastern and western central banking mechanisms would be quite small; for it must be kept in mind that the functions of the Bank Deutscher Laender are very similar to those which, in the Soviet Zone, are carried out by the Deutsche Notenbank. Agreement between the Soviet and western versions could, therefore, be reached along lines transferring to a central bank for all of Germany those powers which in the Soviet Zone and the three western zones are in two central banks. Soviet acceptance of this would be facilitated by the fact that the transfer of these functions would still leave them under state control, since the members of the board of directors are appointed by the government. When the Deutsche Notenbank was founded, the Soviet Military Administration emphasized the similar-

3. The name of the bank was changed in July 1948 to "Deutsche Noten-bank," and it was planned to move the bank to Berlin.

ity between the organization of the Deutsche Notenbank and the Bank Deutscher Laender in order to keep the door open for a uniform system. In spite of this, such coordination would not be very significant so long as the political and economic systems served by the banking systems remain as incompatible as they are at present.

## VII. The Situation in Berlin

Under the occupation, banking in Berlin[4] became subject to the quadripartite control of the Allied Kommandatura. Before and during the war, Berlin had been the financial center of Germany so that its reorganization would have been particularly important. But no reorganization was undertaken; there was only disorganization. In April 1945, all banks were closed by the Soviet Military Administration. After arrival of the other Allies in Berlin, the Allied Kommandatura agreed in July of the same year that previous orders of the Soviet Military Administration were to remain in force as Allied orders. By 1948, the banks were still closed and the accounts blocked.[5]

Banking facilities in Berlin during the period under discussion consisted of three new banks: the Berliner Stadtkontor, municipal savings banks and credit cooperatives.[6] The last two were of minor importance and, in the absence of investment opportunities, did little but administer deposits. The Berliner Stadkontor was established in the summer of 1945 as the only commercial bank in Berlin. The Bank had about 70 branches throughout Berlin, and was authorized to undertake any and all banking business. By 1948, the Bank had no statute, nor had it been provided with capital.[7] In the absence of demand for credit, cash accounted for more than RM 2.2 billion of the total assets of RM 2.5 billion on December 31, 1946. The main

4. The discussion of the situation in Berlin is limited to the period before the complete collapse of the city's economic and financial structure in the fall of 1948.

5. An outpayment of RM 100 was made on deposits at savings banks in December 1947.

6. A fourth bank, the afore-mentioned Garantie- und Kreditbank, also did business in Berlin under Soviet protection, even though its application to do so had been turned down by the Allied Kommandatura.

7. In May 1948, the Magistrat of Berlin made plans for the reorganization of the Bank. Its name was to be changed to "Kreditbank von Berlin" and it was to be provided with a capital of RM 100 million. At the same time, the Magistrat recommended the establishment of a Land Central Bank in Berlin. This reorganization, however, could not be completed, since the Allied Kommandatura, whose approval was required, ceased effective operation at the time. A Land Central Bank was established later for the western sectors of Berlin.

activity of the Bank was the clearing of all interzonal payments between the Soviet Zone and Berlin, on the one hand, and the western zones, on the other. It also served as fiscal agent for the city of Berlin.

With the Berlin banks closed, no repayment of outstanding loans or mortgages was possible. The Soviet Military Administration was opposed to permitting the banks to accept such payments, since this was considered a wedge for an eventual reopening. In April 1947, the Allied Kommandatura agreed to establish a commission consisting of four German delegates, one appointed by each of the four powers, which was put in charge of collecting outstanding loans and mortgages from debtors residing in Berlin. Amounts collected were deposited in the name of each individual bank at the Berliner Stadtkontor and were blocked there. The collection of outstanding debts, however, made little progress. With bank accounts blocked, few debtors were in a position to repay; and since the commission was not authorized to return collateral, those who could repay were frequently unwilling.

In the reorganization of the Berlin banking system, it was the Soviet intention to liquidate the banks and replace them by a socialized banking system. But since in quadripartite discussions no agreement could be reached on their liquidation or reorganization, the Soviet Military Administration undertook a series of steps designed to create such confusion and disorder as to make a reestablishment of these banks impossible.

Already in the last days of the war and the first days of the occupation, the Berlin banks had been severely raided both by Soviet troops and by Germans who used the opportunity of disorder to empty vaults and steal unguarded cash and valuables. In November 1946, the Soviet Military Administration issued an order to remove all securities and documents out of banks located in the Soviet Sector of Berlin; they were to be transferred for "safe-keeping" to a special commission created by the Soviet Military Administration.[8] In the course of 1947, the German Central Administration of Finance for the Soviet Zone was ordered to take over and administer the property of the major banks in the Soviet Sector of Berlin. Their boards of directors were declared out of existence and their functions

8. The Berlin "Wertpapiersammelbank," the most important centre for the collective deposit of a large proportion of securities in Germany, had previously been removed to Potsdam by the Soviets. Repeated Allied protests against these unilateral actions were of no avail. On the contrary, the Allies were informed that these papers would be made available only for the purpose of liquidating the Berlin banks, but for no other reason.

transferred to the German Administration. In a further step, the accounts of a series of banks, including banks located in the western zones of Germany, both at the Berliner Stadkontor and the central post-check office, were closed and blocked.[9] When these actions were protested by the western members of the Allied Kommandatura, the Soviet representative took the novel position that only new banks were within the competence of the Allied Kommandatura, but not the old ones, "and that, in any case, he intended to solve the problem in a radical manner." In the meantime, the banks in the western sectors of Berlin continued to administer their property, though no new banking business was undertaken.[1]

The future of Berlin as Germany's financial center hinges on Allied agreement on a unified Germany. Even before the breakdown of the Allied Control Authority, the outlook for unity had become so dark that many of the remaining bankers in Berlin were making efforts to move to the western zones. The attempt to preserve as much as possible of the old banking structure was slowly recognized as futile. Firm tripartite measures on the part of the Allies in the western sectors of Berlin might have prevented, or at least slowed down, the deteriorating position. But inertia, or perhaps the fear that it might aggravate the split between East and West, prevented the necessary action.

## VIII. Conclusion

The structural reorganization of the German banking system is now completed. In the western zones, the Reichsbank has been put under custodianship pending its liquidation, Land Central Banks have been established to serve as central banks for each Land, and the Bank Deutscher Laender has been created to serve as central bank for the area as a whole and coordinate the activities of the Land Central Banks. The central control of the "Grossbanken" has been broken and decentralized to a Land basis. In the Soviet Zone, banking has been socialized, the old banks "liquidated" and the state made responsible for the central direction of all finances.

The American plan for the reorganization of the German banking system was one phase of the so-called "Morgenthau-Bernstein"

9. These accounts had been used quite legitimately by banks to receive interest payments from customers in Berlin and the Soviet Zone.

1. The Soviet Military Administration protested on various occasions their activities as being contrary to the order closing Berlin banks. But when the Banking Committee of the Allied Control Authority investigated this charge, it was found that the activities of the banks were not only legal, but, in fact, were requested by the Allies, the Berlin Magistrat, and had been specifically authorized in the orders closing the banks.

policy of a decentralized, non-industrial economy for Germany. The question of whether it was at all possible to force on Germany an agricultural economy became academic with the division of the country, at least economically, into East and West: the western zones by themselves certainly cannot become agricultural and still maintain reasonable living standards.[2] While the actual application of the policy was, therefore, modified by the course of events, the underlying assumptions were, in the field of finance, rarely questioned.

Instead, the drift away from the original policy was the result of the necessity to relieve the American tax-payer of the cost of feeding Germany by permitting sufficient industrial output to help Germany pay in exports for minimum import requirements. Minimum living standards had to be assured not only to facilitate the occupation by avoiding disorder and non-cooperation, but were also necessitated by the recognition that in their absence it would be difficult, if not impossible, to interest the German people in the ideals of democracy and keep them from the all too easy acceptance of communism. Moreover, it soon became clear that low industrial production in Germany would not only keep German living standards down, but would also seriously interfere with the economic recovery of Western Europe. These factors soon forced modifications in the earlier emphasis on decartelization, decentralization, and de-industrialization.

Nevertheless, the application of these principles to the German banking system was defended by the American Deputy Military Governor as late as December 1947 on the basis of the important contribution made by the "Grossbanken" to the rearmament of Germany: "Through their political influences and their powerful grip upon the financial and economic life of Germany these banks were able to force independent small German businesses into giant cartels and trusts, and to provide the financial guidance and material aid to the German invading armies which enabled them to exploit the financial and economic resources of the occupied countries." He concluded that this would not have been possible had Germany's banking system been based upon decentralized lines.

It is not quite clear whether decentralization was supposed to serve as a punishment for evil deeds or as a method of making future wars more difficult, or both. It seems that it was not the former, since democratic procedures would have required a trial before punishment. Originally, it was planned to try at least the Dresdner Bank, but the trial was dropped for lack of sufficient evidence; the

2. Even Mr. Morgenthau, in his *Germany is our Problem* (New York and London, 1945), pp. 55–56, assumes for his plan a united Germany.

activities of the Bank were, no doubt, evil, but not necessarily illegal. In addition, it would have been difficult to prove the guilt of the institution beyond the guilt of its individual members.

The policy of decentralizing Germany's banking structure was, therefore, based mainly on the assumption that her centralized system assisted in the conduct of the war in a way in which a decentralized system could not. Not only is it doubtful whether, in a modern state, financial means *per se* could ever be a serious bottleneck to industrial expansion, including armament production, but it seems unlikely that a decentralized banking system would ever be a hindrance to a nation eager to go to war. The measures taken in the western zones could be reversed quickly; but that would not even be necessary. A financial system adequate to do its share in the promotion of sufficient industrial production in Germany to assure minimum living standards is also in a position to provide the financial means for armament production. It is unlikely that Germany's reorganized banking system will contribute to a democratic Germany in a way that the previous system could not have done.

The decentralization of the German banking system will not seriously interfere with the present strenuous efforts towards industrial expansion. Nevertheless, it seems doubtful how long the system will prevail in its present decentralized form. It has not found the approval and cooperation of many German authorities. Officials in Military Government expected this, but hoped that they could enforce it over a lengthy period so that the Germans would, in time, recognize it as workable and accept it as beneficial. However, the eagerness to enforce the original policy in the financial field has weakened considerably with the shift away from it in the industrial sphere. Furthermore, the reorganization is contrary to the trend of events in all other industrialized countries, including the United States. Not only has the independence of local institutions been severely reduced and the control of central institutions strengthened everywhere, but the interference of the state in the financial affairs of the nation has become more direct and extensive. The reorganization of banking in Germany seems now like an unnecessary and futile attempt, the success of which assumes conditions which have not existed for a long time and which are not likely to return in the forseeable future. It is doubtful whether the well-meant attempt has a chance of survival.

HANS A. ADLER.

HARVARD LAW SCHOOL

# THE WELFARE ECONOMICS OF HEINRICH PESCH

SUMMARY

I. Introduction, 342. — II. A goal to be attained: a social science, 343. — III. Objectivity of the goal, 348. — IV. Social philosophy and ethics, 353. — V. A practical and normative science, 357. — VI. Conclusion, 360.

## I. INTRODUCTION

The central issue emerging in this paper is the question of whether the economy has a goal beyond the market-place aims of individuals. The problem arises from the primary purpose of the article, which is to present the nature of economic science according to Heinrich Pesch.[1] To Pesch the nature of the science of economics follows logically from the nature of the economic system. To what extent his position conforms with prevailing views will be indicated by reference to the recent Anglo-American literature on this topic. The significance and applicability of the approach will be shown by a few selected references to his consumption theory.

1. Heinrich Pesch, S. J., was born in Cologne on September 17, 1854 and died at Valkenburg, Holland, on April 1, 1926. His first social studies were made at the University of Bonn in 1872, at which time his professor of economics was Erwin Nasse. On January 10, 1876 he entered the Society of Jesus and was trained during the next fourteen years in classical literature, philosophy, theology, higher mathematics, and the natural sciences. Four years of study in England gave him a realization of the distressed lot of the Lancashire workingmen, one of the main motives inspiring him to dedicate his life to the aid of the laboring class.

After the completion of the formal course of the Jesuit Order, Pesch devoted himself for the next ten years to the private study of economic theories and conditions. Then in 1901 at the age of forty-seven he resumed the formal study of economics. For two years he studied at the University of Berlin under Gustav Schmoller, Max Sering, and the professor to whom he attached himself most closely, Adolph Wagner. The remaining years of his life were dedicated to writing on economic subjects. His major publications are: *Die soziale Befähigung der Kirche* (Berlin, 1891; 3d ed.; 1911); *Liberalismus, Socialismus* und *christliche Gesellschaftsordnung* (2 vols.; Freiburg i. Br., 1896–1899; 2d ed.; 1901); *Ethik und Volkswirtschaft* (Freiburg i. Br., 1918); and his monumental life-work, *Lehrbuch der Nationalökonomie* (5 vols.; Freiburg i. Br., 1905–23; vols. I–III, 2d–6th ed.; 1924–26.) Honorary degrees were bestowed upon him in the closing years of his life by the University of Cologne and the University of Münster.

For further details cf. Pesch's autobiography in *Die Volkswirtschaftlehre der Gegenwart in Selbstdarstellungen*, F. Meiner, ed. (Leipzig, 1924),Vol. I, pp. 190–208; the article, "Heinrich Pesch," by Goetz Briefs in *Encyclopedia of the Social Sciences* (New York, 1937), Vol. XII pp. 91–2; and the brochure, *Heinrich Pesch and His Theory of Solidarism* by Franz H. Mueller (St. Paul, 1941).

Heinrich Pesch was the first theorist who attempted to construct an economic theory based on Aristotelian-Thomistic philosophy. He is sometimes referred to as the leading Catholic economist, which is correct only in so far as Scholastic philosophy is commonly held by Catholics; but its adherents are not limited to Catholics, for it is a philosophic system based on reason rather than on faith.[2]

Adam Smith's classical title, *An Inquiry into the Nature and Causes of the Wealth of Nations*, epitomizes the general approach of Pesch to the study of economics. His concern is the general welfare rather than private profit. He thus defines economics: "The science of the economic life (the process of providing material goods) of a people, considered as a social unit, bound together by the politico-social community life." Or, with direct application to the modern economy: "The science of the economic life of a politically united community, on the level of a developed exchange economy, in relation to the material national welfare as a goal required by the social purpose of the political society" (I, 457).

## II. A GOAL TO BE ATTAINED

Before discussing the problem of the goal of the economy, it should be noted that Pesch, in agreement with most contemporary economists, holds that economics is a science. There is not, however, general agreement about what is meant by "a science." As Pesch points out, the dispute about the nature of economic science, which was a matter of intense interest in Germany at the turn of the century, is only a derivative of the more comprehensive battle in regard to the nature and method of science in general.[3] For him a science in the objective sense is "a summary of the general and certain knowledge and truths extending ultimately to the relatively last cause about a given subject" (I, 461). Some today would rather limit the notion to knowledge acquired by a particular method, such as a priori deduction, or to what can be measured, or can be verified

2. Pesch explicitly asserts that in his work he has "not treated the economic material in a 'theological manner.'" Furthermore it is his belief that though his social philosophic system, solidarism, obtains its complete meaning from a Christian *Weltanschauung*, nevertheless it can be accepted by anyone from the standpoint of pure humanitarianism and the national welfare. (*Lehrbuch der Nationalökonomie*, II (3d ed.; Freiburg i. Br.: Herder, 1920), p. vi.

In the remainder of the article all references to the Lehrbuch will be to the latest editions and will be embodied in the text. The translations from Pesch have been made by the author.

3. Heinrich Pesch, "Eine neue Richtung in der Nationalökonomie," *Stimmen aus Maria-Laach*, LXXX (January 1911), p. 54.

by empirical tests.[4]  In many respects the Peschian concept of economics is closer to the thought of those who deny economics is a science than it is to the position of some who claim the prerogative of a science for economics but attempt to confine it within the narrow borders of the exact sciences.[5]

The social aspect of economics, which is not a matter of dispute among contemporary economists on the formal level of definition, receives particular emphasis in Pesch.  For him economics is not a mere theory of goods.  Goods are needed, produced, exchanged, and consumed by men, who are not Robinson Crusoes, but who live in society.  Moreover, as economics treats of the national whole, it deals with the economy of a politically united people.  Still, he admits it is indisputable that "economic life has its source in human needs, which are given in and for themselves independent of determined and particular social supposits and which clamor for satisfaction in every form of social organization" (I, 502).

The important conclusion which Pesch draws from the social aspect of economics marks the chief difference between the German economist and many modern theorists.  It is that economics, being a social science, is concerned with a goal to be attained: "In the natural sciences it is a question only of the knowledge of causes and effects; in the social sciences, however, to which economics belongs, ultimately about the knowledge of means and results in relation to a desired goal."[6]  And in the *Lehrbuch* he adds the thought that the human conduct and striving, which are the object of research in a social science, are evaluated according to their adaptation to the purposes which they serve (I, 473–4).

4. For an example of the limited concept of science, cf. Lionel Robbins, *An Essay on the Nature and Significance of Economic Science*, (2d ed.; London: Macmillan, 1948), esp. pp. 104–106, 139–41; also his articles, "Live and Dead Issues in the Methodology of Economics," *Economica*, new series, Vol. V, (August 1938), p. 347; and "Interpersonal Comparisons of Utility," *Economic Journal*, XLVIII (December 1938), p. 639.  For a criticism of Robbins in conformity with Pesch's position, cf. Lindley Fraser, *Economic Thought and Language: A Critique of some Fundamental Economic Concepts* (London: Black, 1947), pp. 30–31; and his article, "How Do We Want Economists to Behave?", *Economic Journal*, XLII (December 1932), p. 560.

5. Frank H. Knight: "If we accept the aphorism, 'science is measurement,' as a definition of science, which is its only intelligible meaning, then there is no such thing as 'economic' science . . ." (*American Economic Review*, XXIV, June 1934, p. 236.)  Raymond T. Bye: " . . . Science may almost be said to begin with measurement . . . It is better for the present, to designate economics as 'a branch of learning' than as a science." (*Journal of Political Economy*, XLVII, October 1939, pp. 637–40.)

6. Heinrich Pesch, *Ethik und Volkswirtschaft* (Freiburg i. Br.: Herder, 1918), p. 154.

The goal of the economy is the central notion of Pesch's concept of economic science. This means that "for the complete and correct establishment of the content of economic theory the teleological method is of outstanding importance" (I, 565). For, without first establishing the goal of the economy, there is no criterion of choice between the multitude of ends and means which confront the economist. "Upon the determination of the goal . . . depends the choice of the means" (*Ibid.*). It is this goal which is used as the yardstick to determine which means are favorable, efficacious, or useful. The activities of individuals and organizations, and the production and distribution of goods, are called economically useful in as much as they serve this goal, and only in so far as they serve it.

The antithesis of the Peschian teleological position is to be found in the dictum of Lionel Robbins: "There are no economic ends."[7] It is claimed that "economics is entirely neutral between ends; that, in so far as the achievement of *any* end is dependent on scarce means it is germane to the preoccupations of the economist."[8] If the dictum, "there are no economic ends," is only a cryptic way of saying that without the aid of some other discipline the economist cannot establish the goal of the economy, then, although the words are not happily chosen, there would be no disagreement in thought. If, however, this means that no matter what end is sought, whether to build a house or to live a life of debauchery, whether to produce "guns" or "butter," as long as it is done in an efficient way, it is economics — and this is apparently what Professor Robbins has in mind — then there is a most fundamental difference between Robbins and Pesch. The former considers only the economical use of the means; in fact, one wonders whether it would be a caricature of his position to portray it as "the means justifies the end." Pesch, while giving due weight to the economical application of means, sets up a higher criterion, the purpose for which these means are used economically. The issue has been put most aptly by his follower, Oswald von Nell-Breuning:

> Is there a raison d'être for the economy? In other words, has the economy as a whole, as a material phase of human social life, a *material goal?* Or does just each individual act in isolation, and does it depend only on him which goal he sets up in his economic activity? . . . The question aims much more at this, whether *only* that is true which daily experience teaches us, or whether *beyond* the different goals which individuals choose and *antecedent* to men setting up such material goals in their economic activity, does there simply preexist a *given*

7. Lionel Robbins, *Nature and Significance of Economic Science, op. cit.,* p. 145.

8. *Ibid.*, p. 24. (Italics in original.)

*material goal of the economy?*  This would mean: all activities serving this objective, and producing fruits conducive to it, are exactly *on this account* "economic"; and whatsoever has nothing to do with this objective, has *exactly for this reason* nothing to do with the economy.  All activities aiming at this objective, but producing untoward results, are again *on exactly these grounds* truly "economic," but wrongly applied.[9]

Even from the point of view of efficiency, recognition of the goal of the economy is important, for it is no less wasteful to be "economical" in the pursuit of a false objective than it is to be uneconomical while seeking the proper objective.  As Cannan indicated in his criticism of Robbins, logically, he should have no objection to a man burning down his house to cheat the insurance company, as long as he does not "use more matches than are necessary!"[1]

Heinrich Pesch does not ignore the idea of the efficient use of means.  But he does hold that the principle of economy is broader than economics; it is a principle of reason applicable to all phases of human activity.  "The 'principle of economy' (the optimum proportionality between means and ends, the greatest gain with the least effort) is a general principle of practical reason; a principle of prudent, rational management of affairs, proper not merely to economic life, but extending even to being a cosmic principle" (I, 467).[2] But, while economical choice of means is an indispensable condi-

9. Oswald v. Nell-Breuning, S. J., *Die soziale Enzyklika* (Cologne: Katholische Tat-Verlag, 1932), pp. 61–62.  (Italics in original.)  In general I have followed the English translation of this passage offered in *Reorganization of Social Economy*, B. W. Dempsey, S. J., (ed.) (New York: Bruce, 1936), pp. 84–85, except where I have believed that a more literal translation would help to bring out the emphasis in the original.

1. Edwin Cannan in his review of "An Essay on the Nature and Significance of Economic Science," *Economic Journal* XLII (September 1932), p. 425.

2. In an unsympathetic criticism of Pesch, Abram L. Harris charges "in no place does he completely grasp the notion of economy in its correct meaning of a rational allocation of scarce resources, which constitute a system of means for the satisfaction of wants, the latter being organized into a system of ends."  And he accuses him of a "defective or wrong formulaion of principle" in his view of economy.  (Abram L. Harris, "The Scholastic Revival: The Economics of Heinrich Pesch," *Journal of Political Economy*, LIV, February 1946, p. 41.)

Minor defects in Mr. Harris' criticism of Pesch are that he makes use only of the first edition of the *Lehrbuch I* published in 1905 rather than the latest revised edition of 1924; there is an illogical misinterpretation of Pesch's meaning in at least one sentence quoted; and statements of the principle of economy in a form in keeping with his own are overlooked, for example, in Volume IV, pp. 14, 133, 315.  But the main weakness is that it seems unfair to accuse Pesch of a defective or wrong formulation because he does not always use a form whose acceptance is not undisputed among contemporary economists.  Even Professor Frank H. Knight, the authority to whom Mr. Harris appeals, admits: " 'Cost,' in this sense (in the Marshallian boy-and-berries case) is 'pain cost,' or 'opportunity cost,' as one prefers; there is no real difference in meaning between the two,'' (*Risk, Uncertainty, and Profit*, p. 73.)

tion and cause of most favorable results in economic activity, it still cannot be in itself a final determinant. It is subject to the correct ordering of ends. Technically the question is: What is being maximized?

Moreover, Pesch sees in the principle of economy itself, an implicit teleological principle. The *homo economicus*, who seeks his greatest profit with the least sacrifice, is guided by a goal. He has a very clear idea of what wealth is, and he directs all his efforts toward the end which lies within this concept (I, 469). It is on similar grounds that Lindley M. Fraser has criticized Robbins:

> If he (Robbins) admits — as he clearly does — that it is an *objection* to a given way of achieving a purpose that it is 'uneconomical,' then he is implying that 'economy' is an end and that waste is bad. He is assuming that 'rational,' i.e., economical choice is something worth striving for, and that this end is something with which the economist is directly concerned, and towards which he cannot be indifferent.[3]

Robbins' position that there are no economic ends seems to be the more common opinion today; it is always assumed at the textbook level, and many theorists, omitting all discussion, take it for granted. On the other hand, there are a large number of economists who have explicitly expressed the importance of the study of the goal of the economy. This latter group includes such names as J. M. Clark, Bye, Hawtrey, Wolfe, F. D. Graham, Salz, Walker, and Suranyi-Unger.[4] To quote but one of these, Hawtrey states:

3. Lindley Fraser, "How Do We Want Economists to Behave?" *op. cit.*, p. 568. (Italics in original.)

4. cf. J. M. Clark, "Some Current Cleavages Among Economists," *American Economic Review*, XXXVII, Proceedings (May 1947), p. 9; Raymond T. Bye, "The Scope and Definition of Economics," *Journal of Political Economy*, XLVII (October 1939), p. 627; and his "Some Criteria of Social Economy," *American Economic Review*, XXXIV, Proceedings (March 1944), pp. 1, 7–8; R. G. Hawtrey, "The Need for Faith," *Economic Journal*, LVI (September 1946), pp. 363–64; A. B. Wolfe, "Economy and Democracy," *American Economic Review*, XXXIV (March 1944), p. 3; Frank D. Graham, (Discussion) "Ethics in the Study of Democratic Principles," *American Economic Review*, XXXIV, Proceedings (March 1944), p. 55; Arthur Salz, "The Present Position of Economics," *American Economic Review*, XXXIV, Proceedings (March 1944), pp. 18–19; E. Ronald Walker, *From Economic Theory to Policy* (Chicago: Chicago Press, 1943), passim; Theo Suranyi-Unger, "Facts and Ends in Economics," *Economic Journal*, XLIX (March 1939), pp. 1–13.

It may be well to note that Bye in his "Scope and Definition of Economics," *op. cit.*, on p. 627, after attributing a function to be performed to the economy denies that this is a teleological view, because "there is no implication that the social machinery of economizing has been consciously promoted by any beneficient creator, either human or divine, or that it accomplishes its task efficiently." This denial is based on an erroneous idea of the meaning of "teleological" and fails to appreciate that "function," in the sense of a task to be performed, and "teleological" are correlatives.

Whoever denies that there are rights ends to be sought stultifies all action ... But if economic doctrine is to carry conviction it must proceed from a conception of welfare, in the widest and deepest sense, as well considered as the conduct of men in general. Theory which professes to dispense with that essential requirement is vanity and striving after wind.[5]

It may be helpful to summarize here the main differences on the nature of economic science as between Pesch and the more common contemporary position represented by Robbins.

According to Pesch:

1. Economics is a social science, i.e., a summary of general truths of which we are certain.
2. The economy has a goal to be attained: the providing for the material needs of the people.
3. In economics a teleological approach is required.
4. The formal object of economic science is the goal of the economy.
5. The principle of economy is a subordinate criterion.
6. Economic theory relies on social philosophy and ethics.
7. Economics is a practical science.
8. Economics is a normative science.

According to Robbins:

1. Economics is a social science, i.e., a summary of truths derived from deduction, capable of being empirically tested.
2. The economy has no goal.
3. In economics a positivistic approach is sufficient.
4. The formal object of economic science is alternative choice.
5. The principle of economy is the ultimate criterion.
6. Economic theory is completely divorced from social philosophy and ethics.
7. Economics is a speculative science.
8. Economics is a positive science.

If we look for the fundamental reason for such a series of contradictory propositions, we will find it ultimately in the different concepts of "science" of the two authors. Robbins' scepticism has led him to consider scientific only that which can be measured, or which is subject to empirical test. This faulty epistemology coupled with the desire to build a purely scientific economics has led to his subsequent conclusions.

### III. OBJECTIVITY OF THE GOAL

If the goal of the economy is something more than the sum of individual market-place aims, how does the economist derive a knowledge of it? Is it sufficient to know what the opinion current in a given community at a given time postulates as the goal? Or, is it rather a question of an objective reality?

According to Pesch, unaided economic science cannot determine what is the aim of the economy. Nor is it sufficient to make an empirical study of current opinion, to take prevailing values as data. Knowledge of the goal of the economy is acquired by "philosophic

5. R. G. Hawtrey, *op. cit.*, pp. 363–364.

deduction" from the notion and nature of society and the economy (I, 532–33). Elsewhere he explains that after learning the supreme norm of all social theory from social philosophy, the economist draws the inferences required to achieve the aim of that part of social life, which is called the national economy (I, 476).

On the basis of the social principles of Solidarism, Pesch concludes that the objective, preexisting purpose of the economy is the material welfare of the people, the providing for the material needs of the people. The immediate aim of the economy is to provide the secure social conditions which make possible the general material welfare of the people, the ultimate goal (cf. II, 297–99). The positive expression of the importance of a social order offers a framework for an institutional approach rarely found in economic theory.

Another objective feature of his concept of the aim of economic activity sets Pesch apart even from those economists who postulate a goal, to say nothing of the theorists who prescind from the question. In economic theory today the almost universally adopted position, when speaking about maximization and an optimum, is to refer to the maximization of the *satisfaction* of the consumer. Even when it is purged of all hedonistic elements, and is not a mere pain-pleasure analysis, the study of the process and term-product of the economy is always in terms of subjective satisfaction. This is evident from such statements as Harrod's, "The economic good is thus the preferred"; Hicks's, "Each individual endeavouring to satisfy his tastes"; and Kaldor's, "How much (inequality) secures the maximum of total satisfaction." The subjective level of current theory was brought out most clearly when Professor Robbins, because of his dissatisfaction with satisfaction, popularized the debate on the correctness or incorrectness of interpersonal comparisons of satisfaction. This debate would not have taken place, or been deemed important, had it not been considered an essential assumption of economic theory that everyone has an equal capacity to derive satisfaction. Thus, Maine's Brahmin became of interest to economists because the consumption of goods was discussed in terms of a subjective "utility."

Like Adam Smith, Pesch conceives of economic activity as aiming at something objective, in a physical sense. It is not a question of pain and pleasure, or satisfaction in a psychological sense, but of provision for the material needs of the people. "The physical and psychic enjoyment, which is attached to the satisfaction of an actual want, cannot be the final goal of a struggle so severe and difficult which, for the most part, makes up the content of the lives of men" (IV, 25–6).

If we speak of providing for the needs of the people, it should not therefore be said that economic study must take the path of a psychology of subjective pain and pleasure. Considering the serious purpose of economic theory, that would be nothing more nor less than a theoretical game. No, our study has in view not a pleasure sensation, but the question of the sufficient provision of the people, especially at its broader, lower levels, with good, fair-priced food, clothing, shelter, with all the material goods which they require for the satisfaction of their wants (I, 459).

However, the subjective element in economic activity is not overlooked by Pesch; it plays an essential part in motivation and is one of the elements in value theory. But even here the firm objective basis of a rational satisfaction is supposed:

... evaluation does not rest on mere subjective 'sensation,' or subjective 'feeling,' but on the cognition of something objectively given. It is not an emotion, but a rational act, a judgment, a judgment of value. This judgment has as its object and its cause the objective usefulness of an economic good ... (V, 73).

The significance and applicability of this concept of an objective goal may be observed in his consumption theory. Because of his teleological approach, more than the usual lip-service is paid to the importance of consumption in economic theory. This phase of economic life begins the treatise on the economic process. "The theoretical consideration of consumption theory must be the fundamental law and starting point of economic theory" (IV, 26).

Consumption is defined as "the using up of goods to satisfy a human want" (IV, 25). Reference is primarily to use value, the destruction of exchange value being only a concomitant or consequent phenomenon. For, in consumption we have the ultimate fulfillment of the purpose of the goods-world, when the use value of goods accomplish their function of satisfying human wants.

Perhaps one of the most significant features of Pesch's consumption theory is that the consumption pattern is not a datum. Various criteria are offered with which to judge the suitableness of current consumption.

For example, a relative optimum in consumption is reached only when the standards of hygiene, aesthetics, and morals are observed.[6] Scientific health standards are important in the categories of food, clothing, and housing. At least the minimum requirements of calories

6. Other criteria offered by Pesch are: (1) more urgent needs to be satisfied before the less urgent; (2) a fitting proportion to be maintained between present consumption and the formation of capital; (3) a minimum standard to be attained by all. There would be almost complete agreement between Pesch's position and the criteria offered by Raymond T. Bye in his excellent article, "Some Criteria of Social Economy," *op. cit.*, passim.

specified by the medical profession should be obtainable by all; and the size of rooms and quality of homes should be in accord with modern hygienic studies.  Aesthetic standards are applicable in the matter of style of clothes, the ornamentation and decoration of buildings, the products of the glass, metal, and textile industries.  The principles of morality offer the basis for a distinction between legitimate wants and cupidity.  This safeguards economics from falling into materialism and cynicism; and presents a guide to prevent resources from satisfying irrational perversions at the expense of the true and real wants of the people.  Unless this distinction is observed, the incendiary must be called productive since, in the act of destroying, he at least has his own satisfaction (IV, 147).

This teleological approach has particular significance in the question of "consumers' sovereignty."  It introduces a new note in the discussion, for not only are the rights of the consumer stressed, but also his responsibilities.  True progress in providing for needs is conditioned by the conduct of the consumer — how he forms his demands — how he is accustomed to satisfy his wants.  In reference to the regulation of consumption, in normal times it is substantially a regulation "which the consumer guided by reason and conscience must accomplish for and by himself" (IV, 192).[7]

The idea of an optimum consumption should also be considered in the formation of fiscal policy.  Because taxes and tariffs, especially on food and clothes, influence consumption in varying degrees for the various income groups, a skillful selection and wise moderation in the inevitable taxes on consumption and expenditure will make it possible to lighten in an essential way the burden on the broad masses of the people (IV, 226).  Furthermore, the state can exercise an indirect control over excessive luxuries by taxation laws; although these should not apply to works of quality or cultural accomplishments.

One main objection will certainly be raised against the position of Heinrich Pesch: how objective is the notion of material welfare? Is it possible to establish objective criteria?

Doubt as to the objectivity of the concept of material welfare arises because various schools seem to hold different views on what constitutes the material welfare of a nation.  But this does not prove that the notion is something purely subjective.

It is a question rather about something very real, which indeed in the minds of men, according to their particular personalities, about various subjects can

7. Other regulating factors are the vocational group organizations and the state.

acquire a peculiar coloring. But, on this account, the goal by no means evaporates into the realm of a purely subjective ideal, rather it maintains its completely objective, real value, remains in the realm of objective truth, perceptible by rational cognition and capable of being scientifically established. There are available in reference to it positive and negative valuable criteria, and thus undoubtedly an extensive unity of opinion can be attained (I, 478–9).

If the concept of material welfare were purely subjective, then equally subjective would be the "common good" notion of the legal order, which would mean that political theory represents no reality, and that nothing remains but power politics.

Most, if not all, of the disagreement, in this matter concerns individual cases. But this is not a proof that the general notion is without meaning any more than the failure to agree about the income level at which a man may be classified as poor abolishes poverty.

Even the existence of different schools of social thought does not hinder the reaching of some sort of unanimity about the goal to be achieved. The Mercantilists, Physiocrats, and Adam Smith placed the welfare of the nation in the center of their theories, although they disagreed about how it was to be attained. At least, it should be possible to agree on a negative formulation. A contemporary example is the unanimous agreement that unemployment is an economic evil to be eliminated. And this negative approach can well be the bridge to the recognition of the positive requirements of national economic well-being.

In the matter of establishing objective criteria, no doubt the simplest case is that of hygiene. Medical studies can offer us objective requirements for the health of the nation, such as proper medical care, minimum standards of diet, etc. Also, the demands of morality, though without the same degree of unanimity, are embodied today in various local, state, and national laws. In general, it can be said that, while there are disputed marginal cases, there are also clear-cut cases about which communities can and do agree. In aesthetics, boards of experts are able to agree on certain phases of their field — examples include city planning boards, music and art academies.[8]

It would be well to observe that a quasi-general agreement is desired not because this criterion decides whether a requirement is objectively legitimate, but because without such an agreement the practical application of a requirement may be unattainable. Also, while the emphasis in the establishment of standards is directed to the research and judgment of boards of experts, this in no way means that these boards are to impose their conclusions upon an unwilling

8. For examples of the appeal to scientific studies, cf. E. Walker, *op. cit.*, pp. 255–58; R. Bye, "Some Criteria of Social Economy," *op. cit.*, p. 2.

consumer. Such studies are guides for consumers — as well as for economists and statesmen — who, within legal limits, may exercise freedom of choice.

## IV. Social Philosophy and Ethics

Not specialization but isolation has greatly harmed the modern manner of research. In this wise it could happen, for example, that economic theory became a materialistic and individualistic chrematistic, because its relation with the social sciences was completely forgotten; the wealth of the nation was unnaturally detached from the whole culture of the people, and thus the science deprived of its object (I, 498).

In these emphatic terms Pesch voices his contention that economics is an independent, but not an isolated science. To designate the autonomy of economic science and also the limitations of that autonomy, he makes use of the concept of a formal object, i.e., the particular aspect under which the various factors studied are considered. For the independence of a science is not sufficiently determined by pointing to a peculiar method of study. Autonomy rests on the fact that its subject matter is treated under an aspect different from the aspect under which any other science ex officio considers it. The formal object of economics is the providing for the material needs of the people, the temporal material welfare of the nation.

But an individual is unable to grasp fully the significance of a particular discipline if he does not know its relations with and its dependence on the other sciences concerned with the same subject matter. More specifically, Pesch contends that philosophic knowledge, both social philosophy (as we have already seen) and in the field of epistemology, is necessary for the economist. "The economist must not only assemble, report, classify, and examine, he must also think and draw conclusions, proceeding from both a scientifically established philosophic knowledge and a correspondingly extensive positive erudition" (I, 565). For example, without a dependable philosophy the significance of the regularity observed in social phenomena, which is frequently called an "empirical law," can be misunderstood. It may be supposed that the social events in reality attributable to the motives of free men are rather to be subject to "causal laws" of the natural science type.

In the recent discussions importance of social philosophy for economics has been affirmed by many authors. Fraser characterizes the "pure economist" as "the economist," who is more of an abstrac-

tion than "the economic man" ever was.[9]  Even Professor Hicks admits that beneath Pigou's case of a divergence of the marginal social net product and the marginal private net product lies hidden "some of the gravest philosophic issues about the relationship between the individual and society."[1]  But whatever economists may hold in their methodological discussions, in practice the error has been not so much that philosophy has been avoided, but that a not well-founded philosophic position has been either explicitly or implicitly adopted.  It is the opinion of Arthur Salz that the inadequate philosophy of positivism has become "the official philosophy or the creed of economics."[2]  And, according to J. M. Clark, the cultural lag found in the economic world, where the nineteenth century philosophy of individualism is now combined with group organizations possessing power for good or harm, is "one of the most threatening features of our whole situation."[3]  A defender of utilitarianism, Frank D. Graham, holds: "Economics rests squarely on a cosmopolitan Utilitarianism and there is no way to get it off without a breach of all of the concepts that are the very essence of our discipline."[4]  And Professor Parsons has pointed out that Lionel Robbins' criticism has merely demonstrated the weakness of associating economic theory with an inadequate utilitarianism, but not of associating it with *any* broader conceptual scheme.  He concludes: "What is needed is not the repudiation of the relation of economics to social theory, generally, but its integration with a better and more adequate social theory."[5]

Special consideration is also given by Heinrich Pesch to the question of the relation between economics and ethics or morals.[6]  It is his firm conviction that economic theory cannot be dissociated from ethics or morals.  In as much as the theory of political economy is a practical science (which will be discussed below), it must remain in conformity with the moral law.  Just as economics would hardly

9. Lindley Fraser, "How Do We Want Economists to Behave?" *op. cit.*, p. 557.

1. J. R. Hicks, "The Foundations of Welfare Economics," *Economic Journal*, XLIX (December 1939), p. 707.

2. Arthur Salz, *op. cit.*, p. 17.

3. J. M. Clark, "Educational Functions of Economics after the War," *American Economic Review*, XXXIV, Proceedings (March 1944), p. 64.

4. Frank D. Graham, *op. cit.*, p. 55.

5. T. Parsons, "Reply to Professor Knight," *Canadian Journal of Economics and Political Science*, VI (May 1940), p. 469.

6. Pesch makes a distinction between social philosophy and social ethics. The former determines speculatively the fundamentals of social life, e.g., the authority of the state or the goal of the economy; the latter is concerned with the moral goodness or evil of human acts.

teach the physically impossible, so it should not represent the ethically reprehensible as a scientifically founded truth (I, 503).

His reasons for the necessary and useful harmony between economics and morals are: (1) The moral law has general validity for all times and places. It has reference to the free actions of men; and it is thus applicable to such actions in the field of economics. (2) The moral law directs men's actions to their last end; therefore any contradiction of the moral law is a contradiction between the temporal and eternal end of man. (3) The economist cannot consider the formal object of his science except in relation with the general national welfare. Thus he cannot forget the inner unity of the general culture and of the entire welfare goal of the political society. (4) The material welfare of a people is conditioned by the grade of morality existing in the economic life of the people. (5) The economist can receive from ethics aids for further progress in his theorizing (I, 504–7).

Nor were the many objections raised against this position ignored by Pesch. The two main ones will be discussed here. The first is that this would make economics a part of ethics; it would be a fusion of two disciplines which should be kept separate. He answers: no more than physics by accepting the laws of mathematics becomes a mathematical science. He insists on the proposition, already mentioned, that economics is an independent science, having its own formal object, which is different from the formal object of ethics, which is concerned with what is morally good or morally bad. Even the economic "should be" or "ought" is not an ethical proposal in that it refers directly merely to what should be done in order that the goal of the economy be realized.

It should be noted that this difference of formal objects seems to have been overlooked by some economists, who appreciating the importance of ethics, wish to establish a phase or subdivision of economics to be called "the ethics of political economy" or "ethical economics."[7] This study, it is believed, combines the functions of the moralist and the economist, and can be competently dealt with either by the moralist with economic training, or the economist trained in ethics. This is incorrect. If this study is to be made under

7. Cf. the following statement of Professor Bye: "*Ethical economics* represents the overlapping zone between economics and ethics. It involves the application of standards of right and wrong to the institutions and processes which pure economics describes." (*Journal of Political Economy*, XLVII (October 1939), p. 636.)

Cf. also J. N. Keynes, *Scope and Method of Political Economy* (4th ed.; London: Macmillan, 1930), p. 61.

the aspect of the morality of these economic actions, i.e., whether they are morally good or bad, then it is not an alternative proposition. An inquiry of this kind is the work solely of the moralist who has a knowledge of economics. The economist must be conscious of the morality of economic actions; but this neither requires nor permits him qua economist to *establish* the morality of such actions.

Pesch is insistent on the principle that to accept the findings of ethics, to respect the importance of morals in the field of economics does not require that the economist "become a moral theologian or a preacher of morals, or demand the good as good, or investigate or develop ethical principles and laws" (I, 531). In fact, he warns against an overemphasis of the ethical approach:

> Religion cannot produce corn nor abolish physical evil. Nations with high moral standards will receive economic benefits from the active, particularly social virtues of their citizens, and will better overcome physical evil and difficult times. However, on this account the economist will not then treat the matter of his theory in a theological or moral way, nor certainly should he extract an economic system from Holy Scripture. The medieval scholastics, modern philosophers and theologians treat the facts of economic life under the aspects of morals. That is not the concern of the economist. He will not certainly proceed in opposition to the requirements of moral theology, but at the same time he will not forget that today economic theory has become an independent science, which studies the economic life of a nation under a different aspect than does moral theology.[8]

The second main objection is that economics prescinds in its theory from moral and ethical influences, while conceding them their place in practice. To answer this Pesch discusses the nature of abstraction. He recognizes the place of this methodological device in any scientific investigation, but is opposed to a wrong application of it. Abstraction is not negation. The theorist in abstracting should not change the object being studied, and thus widen the gap between reality and theory. Nor is it enough to claim that non-economic motives are contained in the ceteris paribus of the demand and supply schedules, and then to carry out the analysis solely on the basis of self-interest.

The instinct of self-love as a strong motive in economic transactions is not slighted by Pesch. "The instinct of self-love will always be present, and in objectively morally indifferent matters will always obtain acceptance with most men" (I, 543). But he disapproves of relegating moral considerations to the realm of scientific research of a troublesome incident. The mode and intensity of the economic motive changes not a little according to individual differences and circumstances, and is conditioned by

8. Heinrich Pesch, *Volkswirtschaftslehre der Gegenwart* I, *op. cit.*, p. 202–203.

the finer formation of a sense of honor and by the higher evaluation of ideals. He asks: "Actually, would not every economy and every state collapse, if the 'average' of its members were to permit themselves to be guided only by their own interest, and would wish to deny every consideration except the 'principle of the smallest means?' " (I, 552).

However, Pesch would not deny that "pure theory," as a study of the natural instinct of self-interest alone, has significance for economic theory or possesses scientific worth. His objection is to calling such a study the complete science of economics. With Laveleye, he considers that it is merely the A B C of economic studies, which teaches "what *would* happen" if the economic development were permitted to run its full course under the sole and completely unrestrained operation of the self-love instinct. It is not a complete study of "what is" (I, 553). Self-interest is an instinct, an impulsive force and tendency within our human nature. Only it is not to be forgotten that instinctive forces are subject to man's reason: the guiding law of free rational men belongs to the intellectual and moral order. Personal utility is *a* motive, not *all* motives; and it is a motive, not a norm. To claim that an a-ethic theory is the sole valid economic theory would be a subjective "value judgment" of the highest doubtful value (I, 568).

### V. A PRACTICAL AND NORMATIVE SCIENCE

From the fact that the economy has a purpose Pesch deduces the proposition that economics is a practical science. This is not synonymous with the statement that economics is concerned with reality, which all economists will admit. The question here is whether economic science is satisfied with truth for its own sake or whether it is seeking knowledge in order to achieve a goal. This does not mean that there is no place for pure speculation, but that the ultimate purpose of the science is to implement the goal of the economy.

A common source of confusion is avoided by Pesch, when he distinguishes between the division into theoretical science and practical science and the terms theory and practice. Economics as a practical science is still theory. It includes theoretical elements derived from both speculation and observation; "but because it deals with the aim of the economy, it consequently reaches out to the organization of the economic process in relation to the social goal, and to this extent is a practical science" (I, 474, note).

Most modern economists, except the Robbinsian and the

mathematical schools, consider economics a practical science. For example, to Hawtrey "It is beyond dispute that economic science is concerned with ends and means, and therefore with action"[9]; for Mitchell the importance of economics is derived "from the contribution it may make to welfare"[1]; in Wolfe's opinion, "Economics and political science are *purposive* sciences"[2]; Fraser believes that "in the social studies the end of knowledge is action."[3]  On the other hand, Robbins, by rejecting consideration of the goal of the economy, by declining to permit normative elements in economic science, and by refusing to allow the economist qua economist to give prescriptions of policy, makes economic theory impractical for policy.[4]

Because this practical science is concerned with human conduct, Heinrich Pesch draws his final major conclusion: economics is not only a positive but also a normative science — a controversial proposition — maintained despite the intense criticism directed against it.  Today it is probably the most disputed question in the matter of the scope of economics.

The German economist's argument for the normative character of economics rests on the nature of the object studied:

> If the economic process is not a matter of a 'natural interrelationship', if it for the most part rests on the free actions of citizens, corporations, political judgments, on actions which in no way are 'naturally' determined, which can be as they are or could be otherwise, then it can hardly be understood how economic science properly eliminates the question of an economic 'should be,' how it can renounce the prerogative of being a practical science orientated by a scientifically established and founded norm, and to this extent of being a normative science (I, 475).

In his discussion of the "new method" in economics, Pesch anticipated the objections of Robbins and his school — nor is this surprising for their doctrine is an anglicized version of the position of the German "pure" theorists, especially Max Weber and Richard Strigl.  His criticism manifests a restraint and an appreciation of the commendable in their work.

---

9. R. G. Hawtrey, *op. cit.*, p. 352.

1. Wesley C. Mitchell, (Discussion) "Ethics in the Study of Democratic Politics," *American Economic Review*, XXXIV, Proceedings (March 1944), p. 50.

2. A. B. Wolfe, *op. cit.*, p. 2.  (Italics in original.)

3. Lindley Fraser, "How Do We Want Economists to Behave?" *op. cit.*, p. 570.

4. For a confirmation of this judgment, note the criticism of Fraser: "They (Professor Robbins, Dr. Strigl, and their associates) hold economics to be scientific, not merely in the sense of pursuing objective truth, but in the narrower sense of seeking truth *for its own sake* (rather than for its practical usefulness) . . . (italics and second parenthetical remark in the original) (*Economic Thought and Language, op. cit.*, p. 30.)

We proceed in this study in direct opposition to the new method in economics, which would exclude from economic theory every consideration of an end, every value judgment. If the advocates of this method would only say: 'We will limit ourselves to the study of "what is,"' there would be no reason to object to such a self-limitation. In fact there remains still much to do in economics in regard to the study of causal dependencies. But in most instances they go further: they wipe out from the field of science the problem of 'what should be,' with no consideration of ends, nor value judgments, since this all is unscientific. Moreover it leads to the question of a '*Weltanschauung.*' But there no agreement prevails, and thus in economics, too, unanimity of opinion would be impossible. As if the followers of this method were united among themselves through the study of causes! That these economic writers uphold their position from wholly honest convictions should not be doubted. Too, the scientific productions of these eminent men stand in high esteem. Nevertheless the rejection of the consideration of ends by most of these scholars may be far more influenced by a distorted '*Weltanschauung*' than they perhaps believe.[5]

Though some modern economists approve of the inclusion of both positive and normative propositions in economic science,[6] the more common position appears to be that of Robbins, who believes that there is a logical gulf between the two types of studies "which no ingenuity can disguise and no juxtaposition in space or time bridge over."[7] Propositions with the verb "ought" are just different in kind from those with the verb "is." But this logical gulf has been dug with two faulty shovels, namely, the false assumption that ethics is "unscientific," and the confusion of economic normative propositions with ethics.

Lionel Robbins would not completely reject normative statements, but they are to be relegated to the realm of "unscientific" welfare economics or economic policy. But, this would mean a conflict between theory and policy, a condition which Heinrich Pesch explicitly repudiates: "We cannot agree with the interpretation that wealth, welfare, productivity are the central notions of economic policy, while price is the central notion of economic theory" (I, 476, note 2). Economics as a science is distinct from policy, which is an art. However art and science notionally include each other; for policy, if it is to be carried out completely and effectively, always

5. Heinrich Pesch, *Ethik und Volkswirtschaft, op. cit.*, pp. 152–53.

6. Cf. S. Moos, "Laissez-faire, Planning and Ethics," *Economic Journal*, LV (April 1945), pp. 26–27; R. Hawtrey, *op. cit.*, pp. 362–65; L. Fraser, *Economic Thought and Language, op. cit.*, pp. 36–40, 50; R. Bye, "The Scope and Definition of Economics," *op. cit.*, p. 628; E. Walker, *op. cit.*, pp. 3–5, 210–27.

7. Lionel Robbins, *The Nature and Significance of Economic Science, op. cit.*, p. 148. The treatment of the normative aspects of economics (pp. 147–51) indicates a lack of precision of thought. Throughout there is an unwarranted intermingling and synomous use of the terms "normative" and "ethics"; indeed, Robbins seems to identify all normative studies with ethics.

presupposes a science. The fundamental norm for economic policy is the goal of the economy, which is the central notion of theory.

## VI. CONCLUSION

Heinrich Pesch explicitly offers an economic theory based on a definite social philosophy. Too often it is thought that the dichotomy is whether to accept a "neutral" economic science or one based on social philosophic principles. Actually, this is a false issue. At present there is no absolutely autonomous economic theory; and there is every reason to doubt that a theory of this type could be formulated and still be significant.

In the Peschian concept of the nature of economic science a methodological foundation is presented upon which to construct a theory of welfare, both scientific and significant. Perhaps it is what modern economists are seeking, a welfare economics in the tradition of Adam Smith, Marshall, and Pigou, but with individualism replaced by a sounder social philosophy more in keeping with the needs of our time.

RICHARD E. MULCAHY, S.J.
UNIVERSITY OF SAN FRANCISCO

# CONSTANT PROPORTIONS, FIXED PLANT
## AND THE OPTIMUM CONDITIONS OF PRODUCTION[1]

SUMMARY

I. Introduction: constant proportions and fixed "plant" as restrictions compared to the optimum, 361. — II. The possibility that constant proportions will be a greater restriction than fixed "plant," 362. — III. Three possible locations of constant proportions line, plant line and scale line, 365.

## I

In a recent article in this *Journal*[2] Professor Chamberlin has shown that the common assumption in economic theory of perfect divisibility of factors of production is not to be identified, as it has been, with the denial of economies and diseconomies of scale, i.e. with a linear homogeneous production function and its accompanying constant cost per unit of output. The "long-run" average cost curve, defined to involve complete variability of all factors of production (the "envelope" curve) is typically U-shaped; and instead of some "optimum proportion" of factors being repeated for all outputs, the proportions as well as the total amount of the factors, in general, change with movement along the envelope curve.

If from any point on the envelope curve, $AC$ in Figure I,[3] one factor designated as the "plant" is held fixed while the others are varied,[4] the resulting plant average cost curve (illustrated for different points by the $PAC$ curves of Figure I) is tangent to the envelope curve at the point corresponding to the output for which the plant is optimum. If from any point the *proportions* between the factors are held fixed and their aggregate varied, there results a new curve of "constant proportions" (illustrated for different points by the $CP$ curves of Figure I) tangent to the envelope curve at the point corresponding to the output for which these proportions are optimum. Thus, according to this analysis, "constant proportions" instead of

1. The writer is grateful to Professor Edward H. Chamberlin and to Mr. Herman Myers for discussions of this problem.

2. "Proportionality, Divisibility, and Economies of Scale," this *Journal*, LXII (February 1948), p. 251. The article appears as Appendix B, in the 6th edition of *The Theory of Monopolistic Competition*.

3. This is a reproduction of Figure 4, p. 233, of the previously cited article (Figure 38, p. 233 of *Monopolistic Competition*, 6th ed.).

4. Professor Chamberlin uses the term "plant" to cover, not only a physical plant in the literal sense, but *any* aggregate of factors which is held fixed while others are varied (*op. cit.*, p. 231). Hence a more appropriate term for the cost curve in question seems to be "fixed factor" average cost curve.

being identified with the optimum for all outputs, becomes a restriction, as compared with the envelope curve.

## II

So far so good. But Professor Chamberlin, having categorized constant proportions as a "restriction," still presents it as a restriction less severe than that of a fixed "plant," the *CP* cost curve being described as lying between the plant and envelope curves on either side of their point of common tangency, as is seen in Figure I.[5] It seems that such need not be the case.
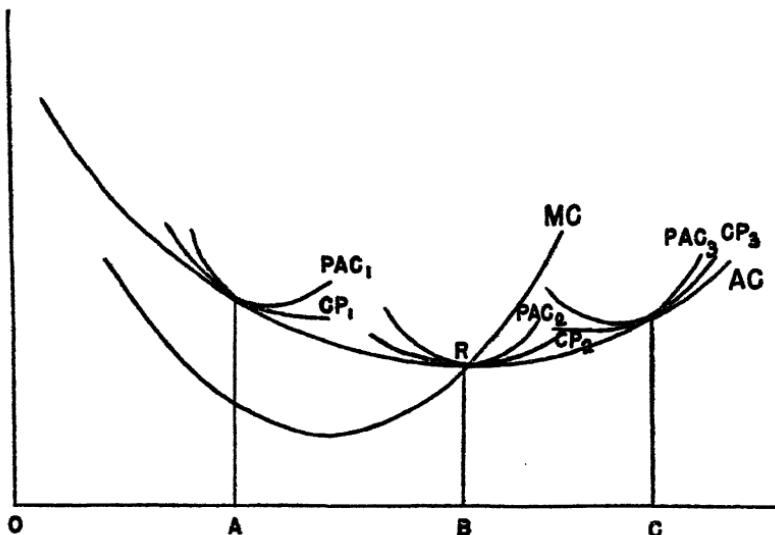


FIGURE I

Let us examine the relations involved as they appear on an isoproduct contour map, as in Figure II.[6] Factors $A$ and $B$ are measured along the horizontal and vertical axes respectively. $I_1$, $I_2$, $I_3$, and $I_4$ are product isoquants, or contour lines; and $A_1 B_1$ and all lines parallel to it, are iso-outlay lines. The point of tangency of a product isoquant with an iso-outlay line is a point of minimum cost for the output in question; and the locus of such points for all outputs is the scale line, $SL$ in Figure II.

If the "fixed" factor is measured along the horizontal axis and

5. *Op. cit.*, p. 251 (p. 249 of *Monopolistic Competition*, 6th ed.).

6. This figure is an adaptation of Figure 6, p. 258, of the previously cited article (Figure 40, p. 256, of *Monopolistic Competition*, 6th ed.).

the variable factor along the vertical axis, then any line perpendicular to the horizontal axis will indicate a certain amount of "fixed input" or "plant" combined with increasing amounts of the variable input as we move up the line; the resulting outputs are given by the magnitudes of the product contours along the way. Such "fixed input" lines are $MK$ and $NL$ in Figure II for amounts of "plant" equal to
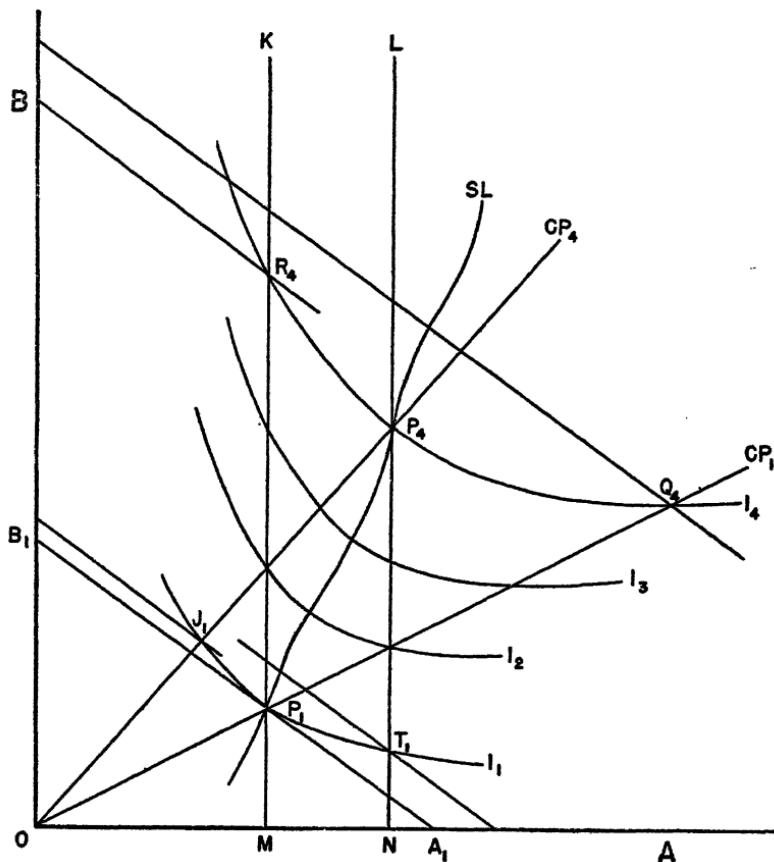


FIGURE II

$OM$ and $ON$ respectively. The "plant" or "fixed input" average cost curve can be derived by dividing the magnitude of each product isoquant into the value of the iso-outlay line which intersects it on the perpendicular for the "plant" in question. This computation gives the average cost of producing all the various outputs with the designated "plant."

Constant proportions of the factors are indicated by a straight line through the origin at a designated angle. Such constant proportions lines are $CP_1$ and $CP_4$ in Figure II, for proportions of $A$ to $B$ equal respectively to $\dfrac{OM}{MP_1}$ and $\dfrac{ON}{NP_4}$. Along each such line the average cost of producing various outputs can be determined in the manner just described for the "plant" average cost curve.

In Figure II the best "plant," $OM$, for output $I_1$, is indicated for all outputs by drawing the perpendicular $MK$ to the plant axis through $P_1$, the point of tangency of $A_1B_1$ and $I_1$, and therefore a point on the scale line, $SL$. The best proportions for this output are indicated by the straight line $CP_1$ drawn from the origin through $P_1$. Once these two straight lines are drawn, it is possible to determine which is the more costly limitation, constant proportions or fixed plant, for outputs other than $I_1$, by comparing the cost of producing the outputs in question under either of the restrictions. In the figure as drawn, the cost of producing output $I_4$ with the proportions of $CP_1$ (the best proportions for $I_1$), exceeds the cost of producing the same output with the plant $OM$ (the best "plant" for $I_1$), since $Q_4$ is on a higher constant outlay line than is $R_4$. Hence, to revert to the average cost diagram, if output $I_1$ in Figure II is taken to be $OA$ in Figure I, and the plant and $CP$ average cost curves are drawn in for this output on the basis of the isoquants of Figure II, these two curves will not take the relative positions shown in Figure I. The reason is that the larger output $I_4$ of Figure II is produced at a lower cost with the plant appropriate to output $I_1$ ( $= OA$ in Figure I) than with the proportions most appropriate for $I_1$. Therefore, to the right of the point of common tangency above $A$ (not lettered in the diagram), $CP_1$ will lie *above* $PAC_1$.

It is evident from Figure II that by changing the curvature of $I_4$ a situation may easily be pictured in which the reverse is true: in which $R_4$ is on a higher equal outlay line than $Q_4$. The $PAC_1$ curve would lie then above the $CP_1$ curve as the two are in fact drawn in Figure I. It is evident, too, that the curvature of $I_4$ might be such that $R_4$ and $Q_4$ would lie on the same equal outlay line, in which case $PAC_1$ and $CP_1$ in Figure I would coincide for the output in question.

Next, let us take output $I_4$ as the point of departure and see how effective the proportions which are optimum for that output will be when employed for other outputs. Along $CP_4$ drawn from the origin through $P_4$, proportions are fixed, and $NL$ defines the "plant" with differing quantities of variable input. The equal outlay line through $T_1$ is higher than that through $J_1$, both of which are points

on $I_1$. Therefore, if output $I_1$ were to be produced alternatively by the best "plant" or the best proportions for $I_4$, the plant restriction would be the more costly of the two. In this case, if output $I_4$ were to be represented in Figure I by $OB$, the curves $CP_2$ and $PAC_2$ would be correctly drawn to the left of $R$ — the latter above the former. But it is again evident that the curvature of $I_1$ might be such that the opposite result would follow; $T_1$ might be on a lower constant outlay line than $I_1$ instead of on a higher one as it appears in Figure II; also, the two points might be on the same equal outlay line.

So far it would appear that in moving away from the optimum conditions of production for any particular output, there is no general presumption as to which will have the more adverse effect upon the cost of the other outputs — fixity of "plant" or fixity of proportions. With perfectly "normal" product isoquants, convex to the origin, anything seems to be possible, depending on simple variation in their curvature. But in our isoproduct contour map, we have examined only the case where the scale line lies between the plant line and constant proportions line. There is much more to be said.

## III

In Figures III and IV the relative restrictiveness of "plant" and of constant proportions is treated in a more generalized manner by associating it with the locations relative to each other of the "plant," constant proportions, and scale lines. Three cases are possible:

    I. Constant proportions line between scale line and plant line.
    II. Scale line between plant line and constant proportions line.
    III. Plant line between constant proportions line and scale line.

In Figure III, the scale line appears as before, lying between $MK$ and the constant proportions line. But it is possible for the production function and the factor prices to be such that the scale line passing through $P_1$ would lie below $CP$ to the right of $P_1$, instead of above it as drawn; in other words to lie between $EG$ and $CP$. This would illustrate our first case of the constant proportions line between scale line and plant line $(MK)$.

However, there is an alternative way to represent it. Let us recall that *either* of the two factors, $A$ and $B$, may be held constant while the other is varied; so that the "fixed" factor, or "plant," may be measured along the vertical instead of along the horizontal, axis. If this is done, $OE$ becomes the "plant" which is optimum to output $I_1$, the horizontal line $EG$ becomes the "plant" line, and the figure as drawn illustrates Case I: that of the $CP$ line between scale line

and plant line (*EG*). Interpreting it in this way, what conclusions
are indicated as to the relative severity of the restriction of fixed
plant as against that of constant proportions?
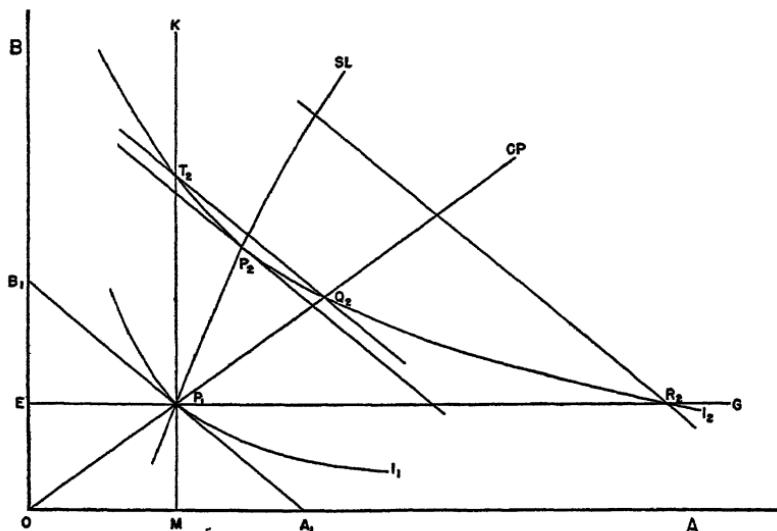


FIGURE III

Let us draw in the product isoquant for output $I_2$, greater than
output $I_1$. It intersects the plant line *EG* at $R_2$ and the *CP* line at
$Q_2$; and the former is on a higher equal outlay line than the latter.
Therefore, for this output the plant average cost curve would lie
above the constant proportions average cost curve, as it does in
Figure I. It is clear that, given isoquants convex to the origin, the
plant restriction will always be more costly when the *CP* line falls
between the plant line and the scale line. With the isoquant for
output $I_2$ tangent to an equal outlay line at $P_2$, there is no adjustment
in the curvature of the convex isoproduct contour which can cause its
intersection with *EG* to take place on a lower equal outlay line than
that which passes through the intersection with the *CP* line. Variations
in curvature may alter the difference, but cannot bring the two to
equality or make the cost higher under the constant proportions than
under the fixed "plant" condition.

The matter may be put in terms of the change in percentage of
total outlay spent on each factor with movement along the scale line.
In the present example the percentage of "plant" outlay to total
outlay is larger along the scale line for outputs in excess of $I_1$ than
it is for $I_1$; and in this case the adverse effect upon cost of holding

"plant" constant would evidently be greater than that of expanding it in proportion to the other factor. For outputs smaller than $I_1$ on the scale line free choice would reduce the percent of total outlay devoted to "plant." Again, the adverse effect upon cost of holding the "plant" constant would be greater than that of diminishing it in proportion along with the other factor.

If we now revert to measuring the "plant" along the horizontal axis and the other factor along the vertical axis, Figure III illustrates Case II — that of the scale line between the plant and $CP$ lines. This is the case already described in relation to Figure II, where we saw that, of the two restrictions, fixed plant or constant proportions, either could be more severe than the other depending on the curvature of the isoproduct contour. In Figure III this contour line for output $I_2$ is drawn in such a way that its intersection, $T_2$, with the plant line $MK$ and its intersection $Q_2$ with the constant proportions line are on the same iso-outlay line. It is thus seen again that by changing the curvature of $I_2$, either $T_2$ or $Q_2$ could be located on a higher constant outlay line than the other. Thus when the scale line lies between the plant line and $CP$ line, the relative position of the three does not in itself give a decisive answer as to which of the two restrictions is more severe.

Again, the conclusions may be put in terms of the change in percentage of total outlay spent on each factor with movement along the scale line. In the case before us, the percentage of "plant" outlay to total outlay is smaller along the scale line for outputs in excess of $I_1$ than it is for $I_1$. Thus since the optimum involves using a smaller *proportion* of plant with both inputs increasing, the two restrictions we are considering — no change at all or a proportionate change in plant — would lie in opposite directions from it. Neither is a more extreme departure from the optimum than the other in the sense that it necessarily lies "further away," and so either one *may* be a greater restriction than the other, or they may be the same.

It should be noted that, if, when any factor, say $B$, is held constant, the relative position of scale line, $CP$ line and plant line make it certain that the "plant" cost curve will lie above the $CP$ cost curve (Figure I), then if the *other* factor — $A$ in our example — is instead held constant, the possibility necessarily appears that the resulting "plant" cost curve *may* lie below the $CP$ cost curve. Therefore this last position of the two curves may not be regarded a mere *curiosum*. And the hitherto common conception of the cost curve of the firm under perfect divisibility of factors as involving the repetition for all outputs of the same (allegedly "best") proportion becomes

strange indeed when it is seen that such a curve may be inferior even to a cost curve derived from holding one factor fixed!

The third of the three possible cases remains to be considered: that of the plant line between scale line and $CP$ line, as illustrated in Figure IV. Output $I_1$ is best produced with $OM$ of factor $A$ and $OE$ of factor $B$, indicated by the point $P_1$. If $OM$ of factor $A$ is now taken as the "plant" to be held constant, $MK$ is the plant line; and it lies between the scale line, $SL$, as drawn, and the $CP$ line.
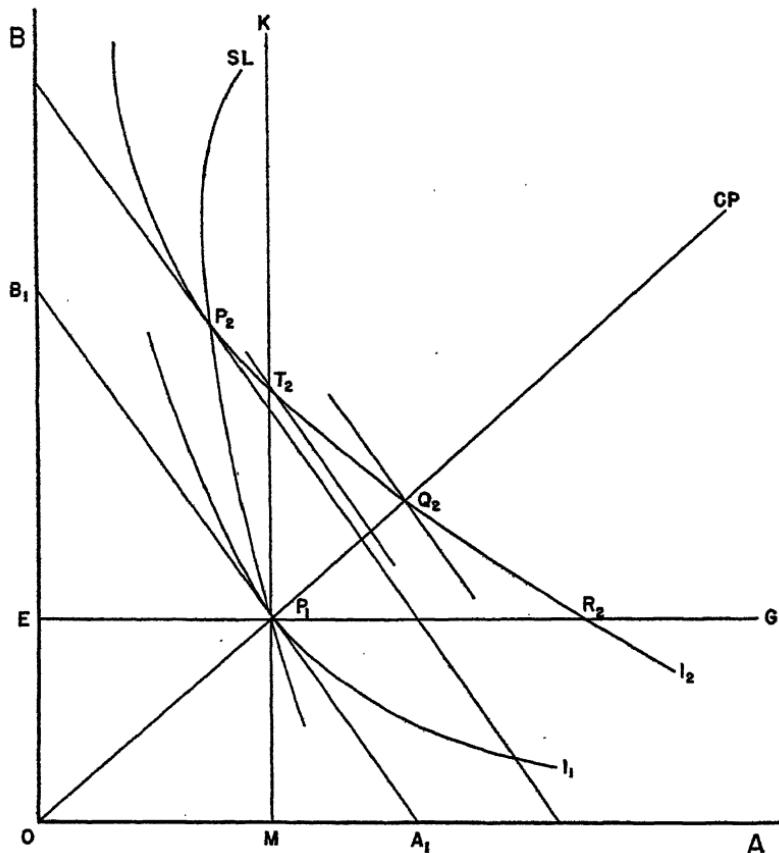


FIGURE IV

Before discussing the likelihood of such a "backward sloping" scale line, let us note that it completes the geometrical possibilities by providing the case where (given isoproduct contours convex to the origin) it is *necessarily* true that the constant proportions restriction is more adverse in its effect on costs than is the fixed plant restric-

tion. The plant and $CP$ lines are drawn with reference to the optimum conditions of production for output $I_1$, and the isoproduct contour for output $I_2$ cuts the former at $T_2$ and the latter at $Q_2$. The equal outlay line passing through $Q_2$ must be higher than the equal outlay line passing through $T_2$. No adjustment in the curvature of product contour $I_2$ consistent with tangency at $P_2$ can make the cost of producing output $I_2$ as expensive along $MK$ as it is along $CP$.

It should be noted that if it is $OE$ of factor $B$ which is held fixed while factor $A$ is varied, $EG$ becomes the "plant" line and we have Case I again, with the $CP$ line between scale line and plant line. The rule already developed for this case applies, since the intersection $R_2$ of $I_2$ with the "plant" line $EG$ is on a higher equal outlay line than is $Q_2$.

What are the possibilities of scale lines as in Figure IV which involve an absolute diminution in the amount of one factor with larger outputs? Clearly, if the factors are defined less broadly so that there are many of them, the case is a very common one: it is a familiar phenomenon that larger outputs may be obtained by technical methods such that certain types of labor or certain machines which are better for smaller outputs may not merely diminish in quantity, but may even drop out entirely. There is no reason to believe that, even in the simplified two-factor analysis, one of the two factors, say "labor," may not be diminished in absolute amount as larger outputs are obtained most effectively by the use of much larger amounts of "capital." It is more difficult to conceive of physical plant (in the literal sense) as diminishing absolutely with larger output. Generally it seems to be true that when there are no limitations on production methods or restrictions in availability of inputs, the amount of durables employed in production increases with output, and the technology which we have acquired tends to confirm that idea.[7] Exceptions, however, are not ruled out. In particular, it is entirely possible that the outlay on physical plant for some smaller amounts of atomic power will be larger than that necessary for larger amounts of power per time unit.[8]

7. See G. F. Bloom, "Note on Hick's Theory of Innovation," *American Economic Review*, XXXVI (March, 1946), pp. 84–85, for some interesting observations concerning this matter. Also John M. Blair, "Does Large-Scale Enterprise Result in Lower Costs?" *American Economic Review*, XXXVIII (May, 1948), pp. 121–153.

8. Indeed the atomic energy example may be too good: *total* costs may be smaller for larger outputs. On page 202 of *The Theory of Price*, Professor George Stigler insists that "the total cost of producing any output must be greater than or equal to the total cost of any smaller output. This is demonstrable in formal terms: if $X_1$ exceeds $X_0$, a firm wishing only $X_0$ would produce $X_1$ and throw

It should be noted finally that the position of the three lines relative to each other may be different at different outputs. Thus the scale line may cross and recross a *CP* or plant line, in which case the *CP* cost curve (Figure I) would be tangent to the envelope curve at several points. Or the scale line may coincide with a *CP* line or with a plant line over a certain range of outputs, in which case the *CP* cost curve or plant cost curve (Figure I) will coincide with the envelope curve over the same range of outputs. Furthermore, even without change in the relative position of the "plant," *CP*, and scale lines, if the latter lies between the other two (Case II, above), "plant" and *CP* average cost curves (Figure I) may cross and recross each other, or coincide for certain distances. The discussion of such phenomena is not carried further in this article.

JOSEPH LERNER.

HARVARD UNIVERSITY

the excess $(X_1 - X_0)$ away." This formal demonstration may crumble before the facts. Professor Stigler is saying that disposal costs are always zero. Actually, they may be negative, zero, or positive. And in some cases they are very important. Clearly there is no "law" that disposal costs must equal zero.

# BILATERAL MONOPOLY
# AND THE COMPETITIVE OUTPUT

## SUMMARY

I. Complications in the comparison, 371. — II. Bilateral monopoly solutions with no income effects, 372. — III. Comparisons with competitive output, 378. — IV. Some assumptions relaxed, 382. — V. Income effects allowed: indifference curves only, 383. — VI. Income effects allowed: indifference curves and production functions, 389. — VII. Conclusion, 391.

## I. COMPLICATIONS IN THE COMPARISON

The concept of "ideal competitive output" of a commodity or service is a general equilibrium concept. Hence, it cannot be dealt with simply or easily. Attempts to compare this output with that under bilateral monopoly are further complicated by (a) implicit assumptions about freedom of entry and the number of individuals or firms; (b) difficulties over the effects of monopolization or atomization of the buying or selling of a commodity on the cost conditions (physical production functions) of the industry; (c) assumptions about differences between firms or individuals that would lead to differential economic rents; (d) the presence or absence of external technical economies or diseconomies which change firm costs as the size of the industry changes even if the prices of the input factors remain constant; (e) the way in which external pecuniary diseconomies (rising prices of inputs) or the falling price of the product as the industry expands, are handled; (f) the way in which the other repercussions through the system are handled to keep everything in equilibrium; and, finally (g) the definition of what we mean by "bilateral monopoly."

We can eliminate the last of these complications by stating at the beginning that we mean by bilateral monopoly the case where a monopolistic seller faces a monopsonistic buyer, but where for purposes of simplicity all other markets are competitive, including the monopolistic seller's purchase of inputs, and the monopsonistic buyer's purchase of other inputs and sale of his product. The reader is warned that earlier writers on bilateral monopoly often meant duopoly, or else the case where a single seller of copper and a single seller of zinc sold to competitive buyers of the two in fixed proportions for the making of brass. For an excellent critical review of practically

everything written before 1943 on all types of bilateral monopoly, the reader should see Henri Denis, *Le Monopole Bilatéral*.[1]

## II. BILATERAL MONOPOLY SOLUTIONS WITH NO INCOME EFFECTS

Comparisons of the price and output under bilateral monopoly with that under competition have usually been made casually by writers whose chief interest was the determination of price and quantity under bilateral monopoly. Professor Fellner, in a recent article in this *Journal*, has come to the conclusion that the maximization of the joint profits of the monopolist and monopsonist in bilateral monopoly where a bargain is made on both price and quantity, *may* result in the same price and *will* result in the same quantity sold as if there were competition on both sides of the market.[2] A simple extension of Joan Robinson's analysis seems to lead to a drastically different conclusion — which appears explicitly in a T. N. E. C. monograph and in two books — while Professor Denis appears to agree with Professor Fellner in one place, and to come to an entirely different intermediate conclusion in another.[3] Furthermore this

1. Henri Denis, *Le Monopole Bilatéral* (Paris: Presses Universitaires de France, 1943).

See also: Alexander Bilimovic, "Der Preis bei beiderseitigem Monopol," *Weltwirtschaftliches Archiv*, LVII (March 1943), 312–363. This is a critical review of the theory, indicating solutions which can be derived if one posits (a) equal gain from exchange, (b) equal negative deviation from absolute maximum gain, (c) maximum collective gain from exchange, or (d) equal absolute satisfaction. The only comparison with competitive output (p. 355) is an indication that the gains of the monopolist can be antisocial if the demand function allows raising the price to the consumers, but this is a result of a pure selling monopoly, not of the bilateral monopoly itself.

See also the article by Tintner cited below. Professor Denis has sought determinate solutions to the bilateral monopoly case using assumptions which seem to us unreal and even inconsistent. For critical comments see Pierre Uri, "Deux Etudes Récentes de Théorie Economique," *Revue d'Economie Politique*, LV (January and February 1945), 71–80.

2. William Fellner, "Prices and Wages under Bilateral Monopoly," this *Journal*, LXI (August 1947), 507.

3. Joan Robinson, *The Economics of Imperfect Competition* (London: Macmillan, 1935), p. 270, and passim.

A. C. Hoffman, *Large Scale Organization in the Food Industries;* T. N. E. C. Monograph No. 35, 76th Congress 3d Session, pp. 161–65; Albert L. Meyers, *Elements of Modern Economics* (3d ed.; Prentice-Hall, 1948), (quoting Hoffman verbatim); William H. Nicholls, *Imperfect Competition within Agricultural Industries;* (Ames, Iowa: Iowa State College Press, 1941), esp. p. 168.

Denis, *op. cit.*, p. 57 n.1, 94–95, and Fig. 26, pp. 96–97. Professor Denis agrees with Professor Fellner that industry marginal cost curve and marginal net productivity curve determine the competitive output where "Des entreprises

problem has been handled by several different, and tricky, graphic techniques. While we shall concern ourselves only with the later and more general solutions, it will be necessary for us to reproduce the bilateral monopoly solutions in order to save the reader the task of rereading earlier and often confusing articles.

To make matters worse, sometimes we are thinking of two groups of individuals with preference systems and fixed amounts of goods (some international trade discussions); sometimes we are thinking about two industries with production functions merely trying to maximize net money profits (Hicks and Henderson)[4]; and sometimes we have a combination (Fellner). Since there are differences between these three cases, it is best to deal with them one at a time. Let us start with the iron ore market case made famous by Professor Hicks. The first thing we must do is indicate clearly our assumptions. In order to eliminate other problems which are not important here, we assume that:

1. The centralization of selling in the one industry and buying of ore in the other will have no effect on the technical production functions or the efficiency of the industry. This assumes away what some writers feel is an important part of the real problem of monopoly in practice, but we are interested purely in the effects of monopolization in a market.

2. There are many ore firms, all alike, and free entry in the competitive case, so that for the industry as a whole there are constant returns to scale, and no differential economic rents. We shall remove this assumption later.

3. There are many steel firms using ore, and free entry, with no rents, as for the ore firms.

4. There is perfect competition elsewhere in the economy, including the purchasing of inputs by the ore industry, the purchase of all inputs except ore by the steel industry, and the sale of steel products. This seems like an unreal assumption, but it will not distort our conclusions about the impact of monopolization on both

concurrentes assuraient d'une part la production du facteur de production, d'autre part la fabrication du produit fini" (p. 97). And he thinks the competitive output is where the marginal cost curve crosses the average net productivity curve of the buying industry where "Des entreprises concurrentes assuraient à la fois la production du facteur de production et celle du produit fini" (p. 96). I am unable to decide just what the distinction is between these two cases.

4. J. R. Hicks, "Annual Survey of Economic Theory: The Theory of Monopoly," *Econometrica*, III (January 1935), pp. 16–18.

A. M. Henderson, "A Further Note on the Problem of Bilateral Monopoly," *Journal of Political Economy*, XLVIII (April 1940), pp. 238–244. One of Professor Henderson's offer curves looks strange to me.

sides of one market.  It would only change the basic offer curves and make them more complicated.

5.  We assume that there are no external *technical* economies or diseconomies in either industry.  If there were, they could be treated exactly as changes in the prices of factors or products are treated here, but again we eliminate them for the sake of simplicity.  The *real* effects on the economy would be different, of course, from those of changing factor prices, but we can analyze them in the same way.

6.  We assume that each of the two industries is large enough so that its expansion *will* affect the prices of factors, and that expansion of the steel industry also affects the price of steel products.  This leads to an asymmetrical situation, for the ore industry will be affected only by rising prices for inputs, whereas the steel industry will be affected both by rising prices for factors other than ore, and by falling prices for steel.

7.  We assume that the size of the net profits in either industry has no effect on its behavior, or on any other prices in the system.  This assumption can be relaxed later.
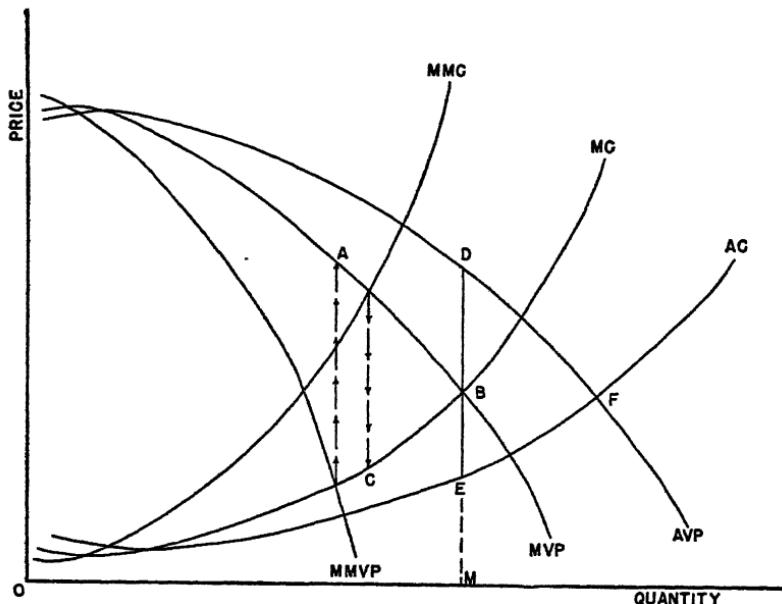


FIGURE I

It is simpler to start out with competition in the ore market, and then centralize the buying and selling of ore.  Since by assumption 7 there are no income effects, we can use a simple type of diagram

on price-quantity axes. Given our assumptions, the average cost for the whole ore industry in producing ore rises for only one reason: the rising prices of the input factors used in mining it. Since the individual firm has no control over the size of the industry and hence no control over the prices of input factors, the industry average cost curve becomes an industry supply curve (*AC* in Figure I). We can picture the typical single ore firm in Figure II.

Similarly, the average net value product of ore used in the steel industry falls because of rising prices of inputs other than ore, and
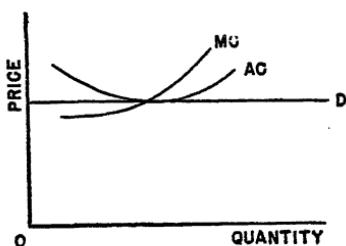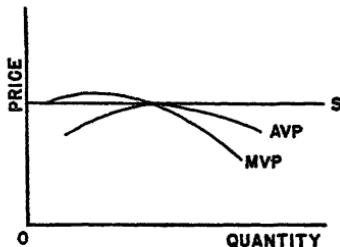


FIGURE II                     FIGURE III

falling prices for steel products. Since individual steel firms have no control over either of these — they depend on the size of the industry — the industry average net value product curve becomes a demand curve for ore (*AVP* in Figure I), and the typical steel firm (they are all alike) is pictured in Figure III. The net value product curves for the *firm* include only technical considerations, input and product prices being given, but for the industry they include also changing input prices and changing product prices. The reader should note that the term net value product is carelessly used here. The *AVP* curve is really total value of the product made in the most efficient way, minus the cost of the other factors, divided by the quantity of ore used. Clearly as the price of ore changes, the proportions of ore and other inputs are changed, through substitution and scale effects, so that this curve is not the average physical productivity of ore times marginal revenue. The average and marginal productivity curves are shifting as the quantity of other inputs changes.[5]

5. It must be added that even this does not make the use of this curve really satisfactory except as a maximum price curve. If the price of ore is pushed below this maximum for that quantity, then the optimum amount of other inputs to be combined with this quantity of ore will decrease. Hence, the whole *AVP* curve will move down. This does not change any of our conclusions, nor narrow the possible range of indeterminacy, though it makes the *MVP* and *MMVP* curves slightly ambiguous.

QUARTERLY JOURNAL OF ECONOMICS

At any rate, the output and price of ore with competition on both sides of the market, are indicated by the point $F$ in Figure I. Now let us put one man in charge of sales for the ore industry, and another man in charge of the steel industry ore purchases, and see what happens. Here we run into the whole gamut of solutions for bilateral monopoly price and output, based on various assumptions, most of them unreal. We shall reproduce the most important ones in order to have some idea of the range of possibilities. We eliminate the whole problem, so important in most real situations, of trusteeship when one person bargains for a group.

A. If the two men bargain for price only, we have a range of solutions, the final result depending upon relative "bargaining power."[6]

(1) If the ore purchaser could set the price, he would presumably attempt to maximize the steel industry profits by equating the marginal cost of ore *to him* to its marginal net value product in his steel industry. However, the amount of ore the ore monopolist will supply at any given price, maximizing *his* profits at this price, is given along the $MC$ curve where this price equals the ore industry's marginal cost (we assumed a rising price for input factors). So the ore industry's marginal cost curve would look like a *supply curve* to the ore-using monopsonist, and we need another curve to show the *marginal* cost of ore to the ore user. This is $MMC$ in Figure I, which bears the same relation to $MC$ that $MC$ bears to $AC$ (not a simple derivative, of course). So the ore using buyer would set the price of ore at $C$ on $MC$ directly below the intersection of $MMC$ and $MVP$.

(2) If the ore producing monopsonist could set the price, he would, in hallowed fashion, equate *his* marginal cost to *his* marginal revenue. His marginal cost curve is clearly $MC$ in Figure I, but what of his marginal revenue? If we look at the ore buyer for the steel industry, we note that, given a fixed price for ore, he will keep purchasing it until ore's marginal net value product in the steel industry threatens to fall below the price of ore. This makes $MVP$ a demand curve for ore, and we must draw in $MMVP$ as a marginal

6. We define bargaining power as Professor Denis does, *op. cit.*, p. 70, as the ability to fool the other person. At the extreme, this would mean the ability to set the best price for yourself and fool the other man into thinking this was your maximum offer. We clearly do not agree with C. E. Lindblom who, seeing the various uses of the term "bargaining power" in the past, insists on including everything under its jurisdiction. ("Bargaining Power in Price and Wage Determination," this *Journal*, LXII (May 1948), 396–417.) Perhaps we should merely drop the term and speak of bluffing power.

revenue curve seen by the ore seller. Again *MMVP* bears the same relation to *MVP* that *MVP* does to *AVP*. The ore seller will then set the price at *A*, directly above the intersection of *his* marginal cost curve (*MC*) with *his* marginal revenue curve (*MMVP*). These solutions are straight from Pigou and Hicks.[7]

(3) It is of course likely that neither man will be able to set the price, each one trying to outwit the other; or even if able to set the price, he may prefer to leave a margin of safety against the risk of failure to reach any agreement. But once the price is agreed upon or set by arbitration or by government, presumably somewhere between *A* and *C*, output (sales) will clearly settle at one spot along the locus *ABC* in Figure I.

B. Others, notably Professors Denis, Fellner, and Henderson, have argued that intelligent bilateral monopolists should not bargain on a price basis only, since it is always possible to find a bargain specifying both price and quantity which will make them *both* better off than under *any* agreement along *ABC* except at *B*, by increasing their joint profits (total profits of the two industries) and giving part of the increase to each. The output of ore (and use of it in making steel) which maximizes the joint profits of the two industries is clearly *OM*, determined by *B* where the marginal cost of producing ore is equal to its marginal net value product (as we have defined it) as an input in making steel. However, at this output, there is a range of possible prices for ore (*DBE*) which will determine the distribution of this profit between the ore sellers and the ore users. We are still assuming here that the profits have no affect on any of the curves. Only along this "contract curve" is it *impossible* to move in such a way as to make both parties better off. From any point off this line, a movement toward it increases the joint profits, and the increase could always be distributed in such a way as to leave both parties better off. The line *DBE* is vertical because of the assumed absence of income (profit) effects.

In practice, of course, it may be quite difficult to make agreements as to both price and quantity, and even theoretically there are serious difficulties. One of the two parties might well refuse to bargain on that basis for fear of ending up in a still worse position in case the other party proved to be a better bluffer. Certainly the ore seller wouldn't let himself be pushed so far toward *E* that he would be worse off than at *C* where he refuses to guarantee any quantity. Likewise the ore user wouldn't let himself be pushed toward *D* so

7. A. C. Pigou, *Principles and Methods of Industrial Peace* (London: Macmillan, 1905), Appendix A.; Hicks, *op. cit.*

far that he would be worse off than at $A$ where he has to pay the producer's price but at least doesn't have to agree to purchase any fixed quantity of ore. The reader himself can set these limits on $DBE$ if he wishes. It would still be to the advantage of one party to *force* the other party to agree on both price and quantity.

C. It is clear that other solutions than $ABC$ or $DBE$ are possible and feasible under quite realistic assumptions about bargaining and bluffing. We do not think, however, that it is likely that any of them would be at much larger outputs than along $DBE$. T. de Scitovszky has suggested a case in international trade, moreover, where the resulting volume of sales is restricted far *below A* or $C$ by a process of successive raisings of tariffs, each time to move along the other country's latest relevant offer curve.[8]

### III. COMPARISON WITH COMPETITIVE OUTPUT

Even if we allow for such general equilibrium considerations as the fact that the changing outputs of each industry, by affecting the prices of factors used in these industries (and of the products as well in the case of the steel industry) might affect in turn the prices of other inputs, or the price of steel products, it is clear that output under bilateral monopoly is considerably less than under competition. How, then, do Professors Fellner and Denis come to the conclusion that $OM$ is the competitive output?

... in the sense that the bilateral monopoly relationship per se does not destroy the competitive pattern of (resource) allocation, provided it does not change the production functions.

If in all other stages of the structure of production, selling and buying were purely competitive (which would imply that the monopsonistic buyer of the raw material was a competitive seller of *his* product), then the allocation of resources would be such as in pure competition.

This output is the "competitive" output, in the sense of being determined by the equality of the marginal cost of the seller with the marginal value product to the buyer. If the buyer-monopolist were a competitive seller, then the quantity of the finished output would be the "ideal" quantity. In this case the marginal value product is at the same time the value of the physical marginal product.[9]

In what sense, or under what assumptions, is $OM$ the "ideal" competitive output of ore? We can think of several cases, i.e., sets of assumptions, which would make the competitive output the same as that when bilateral monopolists bargain for both price and quantity.

8. Tibor de Scitovszky, "A Reconsideration of the Theory of Tariffs," *Review of Economic Studies*, IX, 98–99.

9. Fellner, *op. cit.*, pp. 507, 526; see also Denis, *op. cit.*, pp. 57, n.1., 94–95, 96–97.

1. In the special case pictured in Figure IV, the "contract curve" is narrowed to a point which coincides with the *AC* and *AVP* curves. What has happened here is simply that we have assumed a cost and demand situation such that the maximum joint profit of the two industries combined is zero. Any other output would mean losses to at least one of the industries. In order to have curves tangent like this, we should also have to assume the presence of external economies of some sort, or else the absence of external diseconomies (including rising prices for factors) so that it would not pay to keep restricting output until there was a profit. As Tintner has pointed out, this is clearly a special case, and it is probably *not* what Professors Fellner and Denis have in mind.[1]
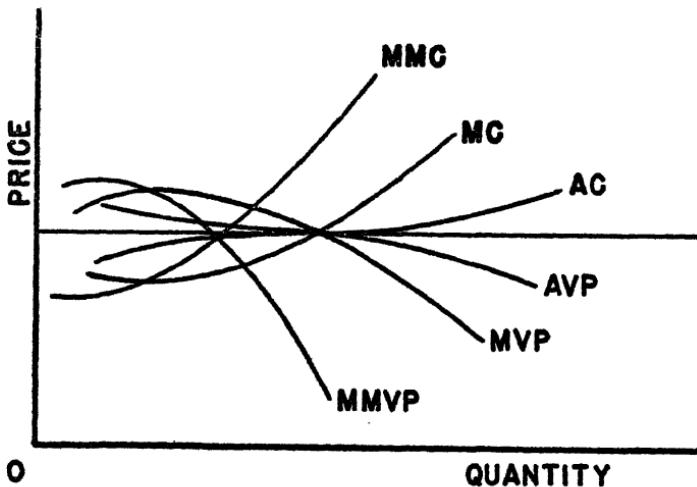


FIGURE IV

2. If we start with the bilateral monopoly situation rather than with the competitive one, it is very easy to make an implicit assumption: lack of free entry of *new* firms — though there may be a large number of firms to start with. In this case, each ore firm would be making profits and would be producing where its *MC* equalled the price of ore. Its *MC* would rise because of diseconomies of scale in a single plant. Each steel firm would also be making profits and would be producing where the marginal net value product of ore to it was equal to the price of ore. If we stick to the assumption that the ore firms are all alike, and the steel firms all alike, *and if we assume*

1. Gerhard Tintner, "Note on Bilateral Monopoly," *Journal of Political Economy*, XLVII (April 1939), 269 and Fig. 6.

*that the prices of all other factors and of steel stay constant,* then we could draw an industry supply curve which would be the lateral sum of all the firm $MC$ curves, and an industry demand curve which would be the lateral sum of all the firm $MVP$ curves. But these same curves are also the monopolist's supply and monopsonist's demand curves once the price is set by someone else. Their intersection is *a* solution then in both the competitive, and the bilateral monopoly case, and if the distribution of the profits has no effect upon output (on the curves) then a whole range of bilateral monopoly solutions (the "contract curve") gives the same output as the competitive case. Note that the $MC$ and $MVP$ curves here are *not* those of Figure I, for our assumptions are different. Here they rise and fall because of diseconomies internal to the firm, and take no account of rising input or product prices. This is a partial equilibrium solution, then, for we assume constant factor prices (except ore) and constant steel prices. Even if we accepted this partial equilibrium solution as satisfactory in some cases, the assumption of no free entry is incompatible with the concept of long run "ideal" competitive output (though not with the concept of short run competitive price, if there are enough firms to start with.)

3. We can, and perhaps this is what Professor Fellner did, make another set of assumptions. Assume again that the prices of other inputs and of steel products are not affected as more ore is used to make more steel, but continue to allow free entry. In this case there would be no equilibrium under our assumption that firms are all alike. We must put in some limiting factor, and we do it by removing the assumptions that firms are all alike: assume that each ore firm after the first must go deeper for its ore and ship it farther, and that each additional steel firm is in a less advantageous location than the one before. In this case the factors affecting the decisions of an individual firm in competition are again exactly those which affect the manager of a monopolized or monopsonized industry if he cannot change the price. Again we can accept the locus $DBE$ in Figure I as a set of possible solutions under bilateral monopoly, provided neither of the two monopolists is smart enough to see that his actual behavior might affect the other's bargaining. The competitive industry supply curve should no longer be labeled $AC$, of course, for it is now the lateral sum of firm marginal cost curves, and similarly the competitive demand for ore is the lateral sum of firm marginal net value product curves. In saying that these are also monopolist and monopsonist supply and demand curves, we implicitly assume that neither has to pay rent to anyone else. Here we have a case where the "contract curve" solution (maximum joint profits) gives the same output as

competition on both sides of the market, the output where $MC$ crosses $MVP$ when both of these are lateral sums of firm curves. Furthermore, this case is compatible with the usual assumptions of competition: large numbers and free entry, both of which are necessary for competitive *output*. It is, however, partial equilibrium analysis, for we have assumed that the prices of other inputs and of steel products have not changed. Once we allow any of these prices to change, or allow external economies or diseconomies of any other sort, the competitive output becomes different from any reasonable bilateral monopoly output. External economies would tend to make the competitive output *smaller*, but it seems legitimate to assume that their influence would in practically every case be outweighed by external technical diseconomies, rising input prices, and falling steel prices. This is a reasonable conclusion, for the competitive output is, almost by definition, the one which maximizes all profits (including consumer surpluses) for society as a whole, not just for two industries. If the monopolist or monopsonist had to pay rent, output would be restricted still further, of course. Implicit assumptions as to who collected rents have been a source of confusion and difficulty in this problem, as in other parts of economic theory.

The confusion, we believe, has resulted from the use of curves which are not relevant for the decisions which are being made, and the use of the concepts of "efficiency" and "ideal output" when we are dealing with partial equilibrium, and a limited segment of the economy. Strangely enough, Professor Fellner carefully points out that the contract curve solution in the wage-employment bargain is not ideal because the preferences of the union may not be those of the labor force as a whole, i.e., ". . . the outcome is optimal *for an aggregate consisting of the firm and the union.*"[2]

2. Fellner, *op. cit.*, p. 528; Another trap for the unwary is Professor Fellner's use of trade union "indifference curves" between employment and the *wage rate*, to illustrate different assumptions about union behavior. These are not really indifference curves of the usual type, but are what Professor Hicks calls "marginal indifference curves" — indifference curves transferred to a price-quantity axis. When Professor Fellner assumes as one possibility, that these "indifference curves" look like regular indifference curves applying to two consumer goods, even as to upward concavity (pp. 511–12), he puts severe restrictions upon the union's preference patterns. Ordinarily there is no reason why *marginal* indifference curves should not slope up to the right, or be convex from above. Indeed Professor Fellner uses the term "concave union indifference map" to mean this very restrictive assumption when he first introduces his third possible assumption (pp. 511–513, especially diagram on p. 513), but in several other places the term means only that the real indifference curves of workers' *income* against employment are not straight lines (see pp. 514–515 where he indicates that all-or-none contracts have advantages for the union only "if the union's indifference map is concave," a statement which only makes sense if he means the income-employment indifference curve, and see also pp. 524–25).

Even when we allow for changing prices of inputs and of steel, however, the analysis we have been giving is not really general equilibrium analysis. There are other repercussions through the system, and we do not feel confident that they can be built into our *MC* and *MVP* curves without circularity of reasoning. It does not seem likely, however, that these repercussions would upset our main conclusions.

### IV. Some Assumptions Relaxed

The question also arises: What if we relax some of the more rigid assumptions? We have already discussed a case where firms were not alike. This raises no real difficulties except perhaps as to the effects of spending money received as rents. If we allow for the effect of profit distribution between the two monopolists, then the curves we have been using become worthless, for they shift with the distribution of the spoils. If we did "build this in" somehow, then our "contract curve" would show a range of outputs, one of which would be the competitive output under the particular assumptions made in 3. above.[3]

As we said earlier, the presence of external technical economies or diseconomies can be discussed for their effect on output, exactly as falling or rising prices of inputs, though the real effects on the economy, and the repercussions, would be quite different.

We have been treating so far only the case where the only motive on either side of the market was the maximization of money profits, with no "income effects" of these profits on the relevant curves. It is necessary because of certain qualitative differences, to make our comparison of competition and bilateral monopoly in two other cases: (a) where only indifference functions are involved (barter), and (b) where indifference sets and production functions are both involved directly. Of course, whenever we start considering successive repercussions and genuine general equilibrium, both types of considerations are involved anyway.

### V. Income Effects Allowed: Indifference Curves Only

Let us take, for our example where only preference sets are involved directly, a hypothetical case of a large number of domestic servants, each with 24 hours a day and no money, and a large number of employers of domestic servants, each with an identical amount of

3. Professor Denis has done this, *op. cit.*, p. 95.

money. We shall start with a more nearly realistic case here, rather than with the simplest one, by allowing the servants to have different tastes and different alternative opportunities for earning a living. We can also allow the employers differences in tastes, provided we insist that all servants be equally effective as servants so that we can have a competitive market. Clearly, since we cannot assume that the distribution of "profits" will not affect the demand or supply of domestic service, we must use a different type of curve where income effects can be built in. We start, as before, with the competitive case, in order to avoid the possibility of making implicit assumptions about freedom of entry. Even if we assume that no one's preferences depend on anyone else's behavior, it is doubtful if we can build up a set of community indifference curves for servants, or a set for employers, since the number of servants, or of employers, is subject to change.

However, given each prospective servant's choice pattern as between leisure and income, and between domestic service and other types of work, plus her alternative job opportunities, we could build up what might be called a "free entry offer curve of domestic help for money at various wage rates." And we could also, given sufficient information, build up another offer curve showing the amount of help that could be secured for any given *average* wage rate with perfect discrimination between individual servants, and perfect discrimination between the different hours worked. This second curve would clearly have no relevance to a competitive market, for it would be impossible to have discrimination there. It does indicate, by its distance from the other offer curve, the surplus, or workers' rent, secured by the servants in a competitive market. We have drawn a hypothetical example of these two curves *OG* and *OF* in Figure V. We measure the total number of hours worked domestically by the distance *down* from *O*, and the total amount of money paid, by the distance *to the right* from *O*. Hence, any point on this diagram specifies a total money sum and a quantity directly, and an average price can be easily calculated. Similar offer curves can be drawn for employers.

The question immediately arises: in a competitive market with free entry or new servants and employers, why should these offer curves bend? This is vital, for depending on the reasons, we shall obtain different conclusions as to the relation between bilateral monopoly and competitive output in this case. With an unlimited number of identical servants with the same alternative opportunities (which we have *not* assumed) these lines would curve only because of something similar to external diseconomies (either technical or

pecuniary) in the two-industry case. For the servants this might be (pecuniary) rising prices of the things they buy with their money because of greater demand or because of fewer people to make them. For the employers it might be falling prices for other things in their budgets as relatively less money is spent on them. These two are not necessarily inconsistent, for the two groups might prefer different types of commodities. We have assumed away "technical" external
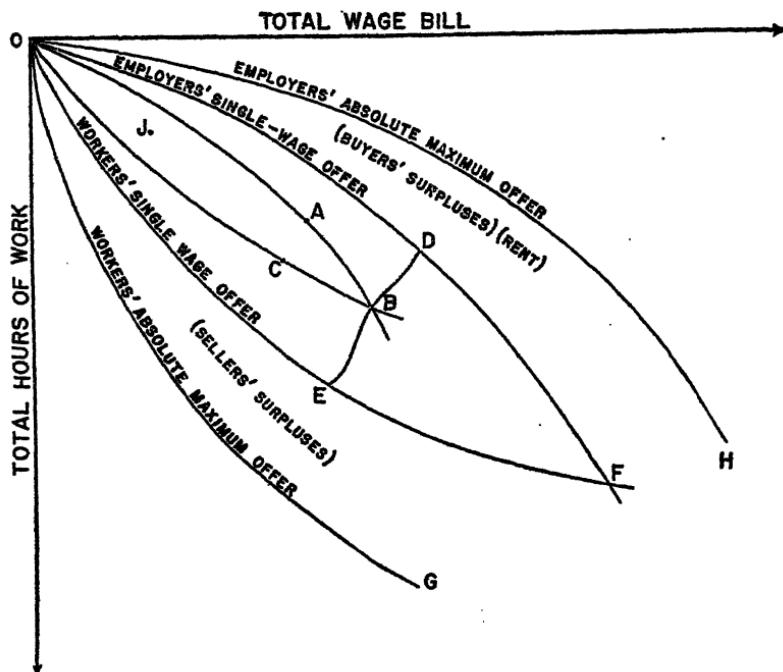


FIGURE V

diseconomies. If they existed, i.e., if employers had less desire for servants as servants became more common in the neighborhood, or servants regarded domestic service as less desirable as more people became servants, we could handle the effects diagrammatically just as we handle the pecuniary diseconomies (rising opportunity costs). The real effects on satisfactions would be different but the comparisons of competitive and bilateral monopoly outputs would be the same. Hence, we disregard this possibility here.

But these offer curves might bend also because of the differences in the preferences of individuals, and in their alternative earning opportunities. (We have assumed that they are all of equal skill as domestic servants.)

In any case there is a *reduction* in the curvature of the single-price offer curve that might otherwise exist, because individual servants will work longer hours as the rate increases, and employers would hire more servant hours each as the wage rate fell. (We eliminate the backward sloping supply curve here for simplicity.)

In Figure V, *ODF* is the offer curve of the employers (assuming no discrimination by servants as between employers, and no reduced overtime charges or other discriminatory devices); and *OH* is their offer curve with perfect discrimination of all types.

Curve *OEF* is similarly the offer curve by the servants of effort for money, given a single uniform wage rate and no discrimination. *OG* is the servants' maximum "no-trade" offer curve under perfect discrimination between servants and perfectly discriminatory incentive systems (higher rates of pay for each successive hour).

Clearly in a competitive market, the *ODF* and *OEF* offer curves are relevant, and *F* is an equilibrium point. What happens now when we centralize the bargaining on both sides of this market? In setting up one employers' representative and one domestic servants' union official, we shall again omit discussion of the questions of trusteeship (vitally important in real life), i.e., problems of whether representatives maximize their own or their constituents' satisfaction, how they know what their constituents want, and how they settle questions of distributive justice within their own group. We assume that the profits of monopoly gained by either side are distributed equally among the members by their (unpaid) representative, and that the impact of these profits on their maximum offer curves could be indicated by building up a set of "indifference curves" with the competitive offer curve as the lowest "indifference curve." These are not really indifference curves, of course. We have already indicated that with free entry (changing numbers) community indifference curves lose their meaning anyway, and further that these curves exclude consumers' or buyers' rents arising from the absence of discrimination. If we were going to allow our monopolists to discriminate, we might want to build a set of "indifference curves" on the absolute maximum offer curves.

Let us start with the case where no discrimination is allowed. What possibilities for profit are there in combined bargaining for either side? In other words, why should there be any curvature in the maximum offer curves of the bargainers? The possibility of "profits" from restriction of amounts of service offered, or used, arises from several sources which must be distinguished from one another:

(a) Why, if employers are willing to restrict their use of domestic

servants, can they get them for a lower wage rate (still assuming no discrimination)? First, as more servants leave for other jobs, those best able to do the other jobs may leave first, so that the alternative job opportunity of the marginal servant (the one about to leave) gradually drops. Second, the wage rates in these alternative jobs may start to drop. Third, the servants remaining may be willing to work fewer hours at a lower rate, although we should probably allow for the possibility of the reverse happening too. Fourth, prices of the things servants purchase may drop, increasing the real value of money relatively to leisure. And finally, fifth, if we allowed it, there might be an external technical diseconomy, domestic service becoming more attractive as less people remained in it.

(b) Why, if servants are willing to work fewer hours, can they get higher average wage rates from their employers, again without any discrimination? First, the marginal employers drop out of the market, leaving all the rest willing to pay more. Second, each remaining employer will presumably be willing to pay more per hour for fewer hours of help, because of the increasing relative importance of service relative to the importance of money. Third, the prices of the other things employers buy with their money may rise. And again, if we allowed interactions between preference systems of the employers, the importance of having servants might increase as fewer people had servants. Here, also, of course, the reverse might happen, reducing the curvature of the offer curve.

Now, we submit, that as soon as either side of the market is monopolized, the bargaining agent will recognize certain of these influences which were outside the consideration of, and not under the control of, the individuals. Hence the offer curve of a monopolist or monopsonist, *even where he cannot set price and has no hope of changing the attitude of those on the other side of the market*, will not be *ODF* or *OEF* in Figure V, but something inside each of these. The individual servants would have been influenced by the first and the third of the items listed under (a) above, and the individual employers by the first and second items under (b), — that is how they got their buyers' and sellers' surpluses — but a bargaining agent for either side would presumably take account of *all* these influences, provided they were important. Hence, with bilateral monopoly, we should have a new pair of maximum offer curves (without discrimination) *OAB* and *OCB* in Figure V.

Again, we can analyze the bargaining possibilities as before, but rather than work through the whole analysis again, we shall merely indicate the solutions with a few comments.

When agreement is reached on price only, or the price is set from outside, we again have a range of possibilities: *ABC*.[4] Clearly it would pay either bargainer to change the wage rate in his favor, even by restricting the offer of money or work, so long as the new bargain was to the left of *OAB* for employers, or to the right of *OCB* for servants. The question again arises why either bargainer would let the other set the price, even by restricting the offer, unless he were either lazy or ignorant of the other's offer curve. For a price once accepted as immutable, we have a locus of outputs *ABC*, the specific price determining the output. However, as we have indicated before, de Scitovszky has shown that successive curtailments of output by each side might put the final solution at *J* in Figure V, where the condition of equilibrium is that each side is facing the other with a sort of "bargaining offer curve" that is tangent to one of the other side's indifference curves. The same criterion is applicable here, though the lack of any real indifference curves in the usual sense leaves the author hesitant about diagramming it.

If the monopolists bargain for both price and total quantity (hours of work), still assuming no discrimination or multipart pricing, we could work out a "contract curve" locus *DBE*, along which their joint "profits" would be maximized. If the bargainers are intelligent, but not too smart and perhaps a bit lazy, these may be likely solutions. Again, as in any bargaining, the attempt to bluff and gain a little more, may give solutions outside this range, though it seems unlikely that the output would ever reach the competitive solution at *F*. The "contract curve" *DBE* is the locus of points of tangency of our "indifference curves" for each side, excluding all buyers' and sellers' surpluses.

What can we say of the possibilities if discrimination of all sorts is also allowed? Presumably solutions inside *OH* and *OG* are possible. Perhaps we could find another "contract curve" maximizing the joint profits of all actual and potential servants and employers, and presumably passing through *F*; but certainly this leaves many difficult questions unanswered.

What can we conclude for this second general case — that of two groups of individuals with preference systems — about the relations between output (quantity sold) under competition as compared with output under bilateral monopoly? It seems clear that so long as there are influences affecting all the individuals in the group, but none of them strongly enough to have any effect, the

4. In all these cases, a public authority could set a price outside the range *ABC*, and one of the offer curves, *OAB* or *OCB*, would still be relevant.

centralization of decision-making for the group is likely to lead to reductions in output. We should hasten to add, that if there are external economies instead of diseconomies (perhaps technical ones, overbalancing the changing opportunity costs) output might be increased, but we think this is the exceptional case. Certainly the problem turns out to be far more complex than previous writers have indicated, and actually we have neglected all but the more obvious of the general equilibrium considerations. Is the competitive output also the ideal one? The answer depends on the type of considerations which influence the monopolist as distinct from the individual in pure competition. If they are external *technical* economies or diseconomies, then we can agree, with Pigou, that maximizing the joint profit of the two monopolists will lead to an output closer to ideal than under competition. On the other hand, if they are pecuniary (changing prices or opportunity costs), then the competitive output is superior. It seems likely that the latter case will predominate.

What if the market is so small that no other prices need be taken into account, and what if we can neglect other interactions outside the immediate market? Here it would seem that, apart from the problem of income effects, Professor Fellner's conclusion might hold; the output might be the same under bilateral monopoly as under competition. But this conclusion should be intuitively obvious anyway, for if there is nothing of which the monopolist takes account which was outside the ken of the competitors, it should be possible to find conditions under which the monopolist would offer just as large a quantity and the other monopolist buy it. It may even be possible that the monopolists are ignorant or lazy and neglect to take account of these "industry effects" anyway. However, it must be remembered that comparisons of competitive and monopoly outputs involve important distinctions between partial and general equilibrium and between immediate and repercussion effects.

## VI. Income Effects Allowed: Indifference Curves and Production Functions

Finally, we come to the third case where indifference patterns and production functions are both involved. It is very easy, as the author has found to his sorrow, to allow examples of this sort to become so complicated that they can be handled only by mathematics, and even then all sorts of index number problems arise. On the other hand, it is also possible to simplify the problem so much that it becomes meaningless. As an instance of this, if we insisted on

identical individuals and identical firms, and free entry, and no external economies or diseconomies, all offer curves would be straight lines, and there would be no equilibrium. Or if we posit a very simple economy where all workers are employed either in making the product under consideration, or in the industry using that product (perhaps consumption), then the repercussions become extremely important. If we reduced output in the two industries, where would all the workers go? And what would happen to the demand for the final product? We cannot in this article list all the possible sets of assumptions and discuss them, though a longer treatment might well designate them, as Edgeworth once did, by large and small letters and give results for each possible set.

Let us then take the following case: a large number of small wheat farms, and free entry of new ones, but each additional farm on poorer land so that there is rent. Assume also no lumpy factors so that each farm finds the marginal value product of labor falling as more men are hired. The only cost in raising wheat, aside from rent, is labor cost, and there are a large number of available workers, all equally efficient and with identical taste patterns. For the time being let us assume that the supply can be increased or decreased by bidding the workers from other jobs, with the usual increase in the marginal worker's alternative opportunity because of rising wages elsewhere. Notice that we have removed a restrictive assumption on the firms — that of identical costs, and have put one on the individuals — identity of tastes. The reader should be able to figure out how to handle other variants.

Let us first diagram the competitive solution. We have a market for agricultural labor here, a very abstract market far removed even from the dynamic problems discussed by Alex J. Morin in a recent article.[5] Let us use the same Edgeworth-type diagram again. In Figure V we measure to the right from $O$ the amount of money paid in agricultural wages, and down from $O$ the number of hours (per year) worked. Although we can consider the workers' offer curve as built up from a sort of community indifference map, it is a weird sort of community, with a varying number of people. Again, we can draw $OG$, the maximum amount of work that could be secured on farms by a perfectly discriminatory system of payments (overtime premiums etc.), although here we have eliminated by assumption any differences between the workers' tastes or their efficiency in any

5. A. J. Morin, "A Note on Bilateral Monopoly with Special Reference to Seasonal Agricultural Labor." *Journal of Farm Economics*, XXXI (February 1949), 101–115.

possible job, so that no discrimination between individuals would be possible. Then *OEF* becomes the offer curve of farm work for money, given a single hourly wage rate. The distance between these two offer curves here represents only one type of seller's surplus, which we might call the "usual" type.

As for the farms, which are run automatically, we have one maximum offer curve, *OH*, showing the amount of money that could be got by the workers if they discriminated between farms so as to take all the rents as wages too, and we have the competitive offer curve, showing the offer of money by farms with a single uniform wage rate. The distance between these curves obviously represents differential land rent.

The competitive solution is clearly at *F*, the farms making no profits above their differential land rents (because of our assumption of free entry). Now let us put one union leader in charge of farm workers, and one labor relations expert in charge of hiring farm labor. The question, as before, is clearly whether there is anything that would cause them to restrict the amount of labor supplied, or purchased, below the amount indicated by the offer curves *ODF* and *OEF*, even if the price were set by the other side, or by the government. What sort of things might do this? The answer is that increased output of wheat might cause its price to fall; or increased earnings of workers and their movement to farms might cause prices of other things (besides wheat) to rise. We might also have interactions between preference systems, and between the farm costs systems, e.g., less wheat rust when less is produced in an area. It should be clear that the reactions that must be considered are complex once we get beyond the first stage, and that they may be indeterminate: the price of wheat may tend to rise because more workers can buy more bread, and to fall because more workers are producing more wheat. It seems likely however, that again there will be another set of maximum monopoly offer curves *OAB* and *OCB* with the usual possible equilibrium points.

## VII. Conclusion

We have spent so much time on this matter not simply because it seemed worth-while to clear up the reasons for the disturbing · discrepancy in conclusions reached by different theorists, but also because we feel that the bilateral monopoly type of market is likely to become more and more important, and that its consequences for output and prices relative to competition need to be discussed.

Whether one monopolist maximizes his profits and the other casually makes the best of it, or whether they get together and maximize their joint profits, it seems clear that in all but the exceptional cases, output will be less than under competition, though it may well be greater than under monopoly on one side and pure competition on the other. It seems intuitively obvious that the output which maximizes the joint profits of two groups in society will not usually be the same as that which maximizes the profits and surpluses of society as a whole (the competitive output).

<div align="right">JAMES N. MORGAN.</div>

BROWN UNIVERSITY

# INTERREGIONAL PAYMENTS COMPARED WITH INTERNATIONAL PAYMENTS

## SUMMARY

I. Interregional trade in the United States, 392. — II. Traditional explanations, 393. — III. Payments data for the United States, 396.— IV. The role of Treasury transfers, 400. — V. Summary, 407.

Why in the past has the international gold standard collapsed while at the same time internal gold standards in European and American nations functioned without apparent strain? This question has been asked frequently, and answered in different ways by various economists. The present article is an attempt to evaluate these a priori answers in the light of the intranational experience of the United States during the trying thirties when international payments were so difficult of settlement.

## I. INTERREGIONAL TRADE IN THE UNITED STATES

Trade between different sections of the United States is conducted on the basis of a monetary mechanism which closely approaches a full and complete gold standard with a minimum of impediments to trade. The Federal Reserve System is so organized that the country is divided into twelve partially autonomous regions, each with centralized bank reserves of gold, interrelated by a system of par clearance with virtually invariant exchange rates and free gold movements. The reserves of each region are determined in large part by its commercial and financial transactions with other regions, by its trade, that is, with the rest of the country and with the world. When the receipts of any one region in interregional trade are in excess of its payments to all others, the difference is settled by means of a gold earmarking operation in favor of this region on the books of the Interdistrict Settlement Fund in Washington. This gold movement is akin to the flow of specie internationally; the bank reserves of the paying regions are decreased; those of the receiving region increased. Any region, then, with a chronically passive balance of payments on commercial and financial account would suffer a continuous loss of bank reserves. What has prevented such a continuous loss or gain on the part of any single region, especially during periods of widespread economic maladjustment?

Interpreted most broadly such a question is concerned with the differences between international and intranational trade; more

narrowly the problem resolves itself into the query: "Why does one pay so much attention to international, and no attention whatever to interregional balances of payments?"[1] The breakdown of international trade in the 1930's was the result of disequilibria in international balances of payments; why were there no such difficulties in settling intranational balances?[2]

## II. TRADITIONAL EXPLANATIONS

The explanations offered have varied in scope as widely as the statement of the problem. Those most frequently cited and most generally accepted as responsible for differences in the operation of a gold standard internally and internationally are (1) the fact of a common or unified currency in use within a country, whereas different currencies are used by different countries; (2) the fact of a homogeneous economic policy (monetary, fiscal, trade and migration policies) governing all regions within a nation, while at the same time different nations pursue independent and divergent policies; and (3) the explanation of the classical economists, the fact that factors of production move freely within a country, but are relatively immobile as between countries. It will be shown that the most important determinant in the maintenance of regional balance-of-payments equilibria in this country has been the mobility of productive factors, especially that of capital.

The first of these explanations can be disposed of quickly. While it is true that the regions of the United States are joined by a common currency, this fact in itself is not sufficient to insure equilibrium in the balance of interregional payments. The payments and receipts of any region or country are determined by the productivity of its economy as compared with that of other regions, together with the level and nature of demand. If a region's balance of payments should become passive because of a decline in demand for the products of its export industries, the fact that it uses the same currency as do other regions would not prevent it from losing reserves.

1. M. A. Heilperin, *International Monetary Economics* (London, 1939), p. 9.
2. Although the event was never described in terms of a balance-of-payments disequilibrium, the bank-closings in the early thirties, which originated in agricultural sections of the west and then spread over the country, indicated the existence of intranational disequilibrium. The suspension of payments on the part of banks in any one region was as much an abandonment of the internal gold standard as were the more obvious steps of controlling gold exports taken by England and other nations during the same period. Actually balance-of-payments difficulties were existent in this country in 1932–33, although few people were aware of them as such.

Similarly, even if all countries of the world used the same currency, one nation would suffer a persistent loss of reserves if international supply and demand conditions should change against it.

Actually, the only difference between a unified world currency and an international gold standard lies in the fact that the former insures invariant exchange rates whereas the latter allows exchange rates to vary within very narrow limits. While it is true that variations of exchange rates within the gold points can and do set in motion short-term capital movements that would not otherwise occur, this particular type of capital movement is not disequilibrating to the balance of payments. For all practical purposes, so long as there is free convertibility of one currency into another, the gold standard supplies a unified currency, but this fact is not sufficient to ensure international equilibrium.

For the same reasons a system of international par clearance would not prevent the cumulative loss or gain of reserves by any single country; rather it would offer a more facile mechanism for the accomplishment of the flow. The existence of a clearing fund to maintain international clearance at par would not have prevented the collapse of world trade in the early 1930's. At this time agricultural countries experienced a loss of reserves because of a passive balance of payments. The latter was due to a decline in the quantity and price of their exports together with an inelastic demand for imports. Under such conditions any international agreement for par clearance would have been abandoned by these countries, as was the gold standard, because of the exhaustion of their credit in the clearing fund. Agricultural countries were forced to abandon the gold standard because of their adverse balance of payments position. Neither a unified world currency nor a system of par clearance would have prevented their payments from exceeding their receipts.

Would, however, a homogeneous economic policy governing the entire world at this time, have prevented the balance-of-payments problems from arising, and thereby served to maintain the successful operation of the gold standard? The causes of the adverse balance of payments position of the agricultural economies were not confined to southern Europe and South America. They also characterized the interregional economic position of the agricultural sections of the United States. In this country as elsewhere the demand for agricultural products suffered a sharp decline. Here, too, the demand of these regions for industrial goods for consumption was relatively inelastic, and yet no observable balance-of-payments problems arose, especially after 1933. Can the successful operation of the

internal gold standard here be explained in terms of interregionally homogeneous economic policies?

Actually not. Homogeneous economic policies imply that every region of the country is subject to the same policy decisions and regulations in the monetary, fiscal, trade and migration fields. In order that these policies be homogeneous, it is not sufficient that each region be subject to the same policy-making authority, which is the case in the United States. What is important is that any single policy decision of a policy-forming body should apply to all regions without discrimination. The latter is not true in the United States even on the national, as opposed to the state, level.

Among the many policy-making agencies of the federal government, there are only a few which divide the country into sections or regions for administrative purposes. Most apply their proposals on a nation-wide scale without any apparent differentiation among regions. Insofar, however, as these boards deal with one specific industry or type of economic activity which is highly localized geographically, the effects of administering the policy result in regional discrimination. For example, a good part of the regulations of the Securities and Exchange Commission apply implicitly or explicitly to the New York stock exchanges alone. Changes in these policies will affect the economy of the New York Federal Reserve District, and thereby its interregional balance of payments to a much greater degree than any other. The same would be true of changes in minimum margin requirements made by the Federal Reserve Board. The policy is of national application, but its effects fall most heavily on the New York district. Again, agricultural policy aimed at supporting the price of a certain crop will affect the regions producing this crop in one way, those buying it in another. Thus, in some instances these regional differences in policy results are a matter of degree; in other cases, however, the difference is one of kind.

Many economists have stated as the chief reason for the difference between international and internal trade the existence of different monetary policies among nations, the fact that a flow of gold from one country to another can cause a contraction of a certain amount in the credit structure of the losing country, but an expansion of a different amount in the credit of the receiving country because of different reserve requirements. But this same phenomenon is equally true in the case of the United States. The loss of reserves by the Boston to the New York Federal Reserve District, for example, can result in a greater contraction in the former, smaller expansion in the latter since usually reserve requirements for banks in Central

Reserve cities are greater than in Reserve cities, and the required reserves of the latter exceed those of country banks. Thus the Federal Reserve Board, in requiring different reserves in different regions, does not practice a homogeneous monetary policy.

In regulating discount rates, too, the Federal Reserve Board pursues a non-homogeneous regional policy. Even if discount rates throughout the country are all changed to the same degree, there may be varying effects on regional balances of payments, depending on the level of the marginal efficiency of capital within the different districts. Likewise, in determining the participation of each district in open market operations, the policy of the board results in greater effects on the payments of some regions.

Another agency which divides the country into regions for administrative purposes is the Interstate Commerce Commission, which, for example, allows different freight rates to be charged in different territories for transporting the same commodity by rail. In such a case, however, once the regional economies have adjusted themselves to a given level of rates, it is only a variation in the relative regional levels (such as has recently been effected) which is disturbing to the balance of payments, not the existence of the different levels.

Enough examples have been cited to indicate that regional equilibrium in the United States cannot be explained in terms of homogeneous economic policy, even on the national level. When we take into account also the differences in state policies, and the hindrances to interstate movements of goods and capital[3] which exist, the similarity to the conditions of international trade becomes even clearer.

### III. Payments Data for the United States

Let us now consider the existing data relating to the volume of trade among the twelve Federal Reserve districts of the United States. The amounts of gold reserves gained or lost annually by each Federal Reserve district are shown in the annual reports of transactions through the Interdistrict Settlement Fund. The annual summary of these transactions reveals the balance of payments of each region for that year: the excess of receipts or payments in interregional trade. These transactions are differentiated by the Fund into three categories: transit clearings, Federal Reserve note clearings, and inter-reserve bank transfers.

3. See, for example, W. Adams, and L. E. Traywick (eds.), *Readings in Economics* (New York, 1948), pp. 172–173.

The first group, transit clearings, represents all transactions undertaken by individuals and institutions (with the exception of the federal government) in any one region with all outsiders. Checks drawn on and by member banks and their correspondents are cleared through this account. Although it is true that a sizeable volume of checks is still cleared outside of the Federal Reserve System through correspondent banks in other sections, payments and receipts in this correspondent clearing system must in the long run be in balance. If they are not, the bank whose deposit is being increased will eventually transfer it home via the Interdistrict Settlement Fund. It is, therefore, safe to assume that the net debit or credit of any Reserve Bank in the I. S. F. represents the excess of payments or receipts in all interregional transactions, with the exception of currency movements and transfers of the federal government. The transit clearing account includes the settlement of various inter-reserve bank trans-

NET TRANSIT CLEARINGS AND INTER-RESERVE BANK TRANSFERS
OF THE BOSTON, NEW YORK AND MINNEAPOLIS FEDERAL RESERVE
DISTRICTS, 1919–1939[1] (millions of dollars)

| | BOSTON | | NEW YORK | | MINNEAPOLIS | |
|---|---|---|---|---|---|---|
| | Net Transit Clearings | Net Inter-bank Transfers | Net Transit Clearings | Net Inter-bank Transfers | Net Transit Clearings | Net Inter-bank Transfers |
| 1919....... | +509 | −512 | −3,030 | +2,905 | +279 | −325 |
| 1920....... | +567 | −492 | −565 | +248 | −96 | +90 |
| 1921....... | +310 | −224 | −1,043 | +745 | +92 | −63 |
| 1922....... | +154 | −177 | −576 | +414 | +43 | −21 |
| 1923....... | +158 | −103 | −132 | +83 | +11 | 0 |
| 1924....... | +66 | −49 | −337 | +407 | +39 | −17 |
| 1925....... | +92 | −147 | −320 | +539 | +1 | −10 |
| 1926....... | +220 | −155 | −591 | +626 | +29 | −14 |
| 1927....... | +243 | −203 | −399 | +577 | −38 | +19 |
| 1928....... | +125 | −136 | −240 | +446 | −19 | +18 |
| 1929....... | +95 | −50 | −153 | +359 | −38 | +57 |
| 1930....... | +142 | −145 | +129 | +314 | −25 | +18 |
| 1931....... | +146 | −129 | +590 | −211 | −46 | +31 |
| 1932....... | +78 | −80 | +1,226 | −819 | −116 | +81 |
| 1933....... | +183 | −171 | +436 | −371 | −22 | +42 |
| 1934....... | +248 | −132 | +957 | −856 | −267 | +226 |
| 1935....... | +193 | −86 | +1,270 | −718 | −269 | +156 |
| 1936....... | +155 | −123 | +714 | −906 | −277 | +179 |
| 1937....... | +178 | −214 | −421 | +250 | −183 | +202 |
| 1938....... | +259 | −63 | +382 | −390 | −160 | +134 |
| 1939....... | +293 | −80 | −275 | +545 | −172 | +45 |

[1] + sign indicates net credit (receipts); − sign indicates net debit (payments). Source: Federal Reserve Board, Annual Reports.

actions; open market operations of the system affect this account, as does the allocation of foreign central bank deposits with the Federal Reserve System among its 12 individual banks.

The second group, Federal Reserve note clearings, represents interregional currency movements. It shows the volume of notes issued by one bank which are presented for redemption by all others.

The net debit or credit of any bank on these two accounts (transit and note clearings), taken together, shows the net loss or gain of gold by that region which results from its "foreign" trade. If a balance of payments were to be constructed for this district, its payments for commodities, services, long-and short-term loans would exceed its receipts (or vice versa) by this amount.

The third group, inter-reserve bank transfers, reflects changes in the deposit of the federal government with each Federal Reserve

NET TRANSIT CLEARINGS AND TREASURY TRANSFERS
OF THE NEW YORK, BOSTON AND MINNEAPOLIS
FEDERAL RESERVE DISTRICTS, 1929–1939.

(In millions of dollars)

NEW YORK

BOSTON

INFLOW

$300

200

TRANSIT
CLEARINGS

100

0

-100

-200

TREASURY
TRANSFERS

OUTFLOW

1929    1931    1933    1935    1937    1939



MINNEAPOLIS

INFLOW

$300

200

TREASURY
TRANSFERS

100

0

-100

-200

TRANSIT
CLEARINGS

-300

OUTFLOW

1929    1931    1933    1935    1937    1939
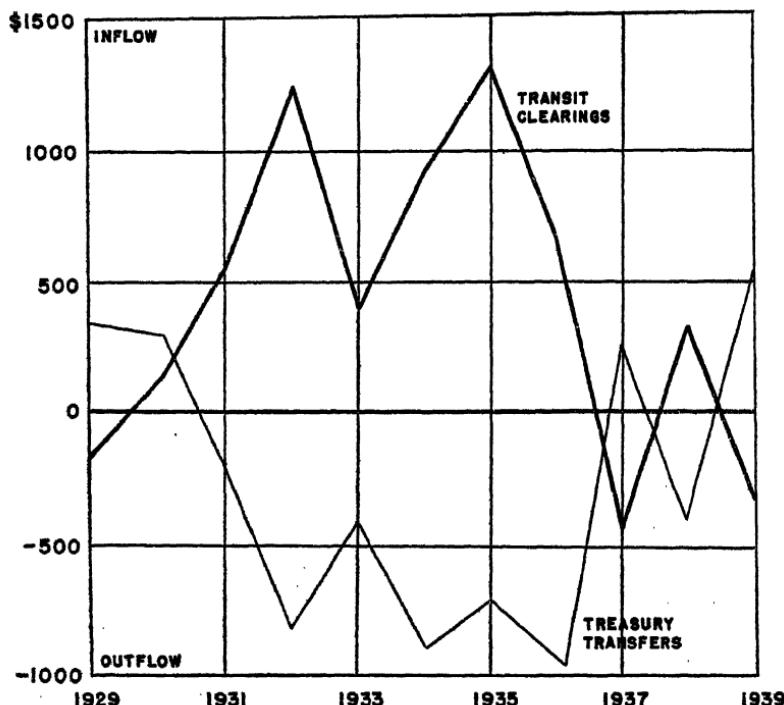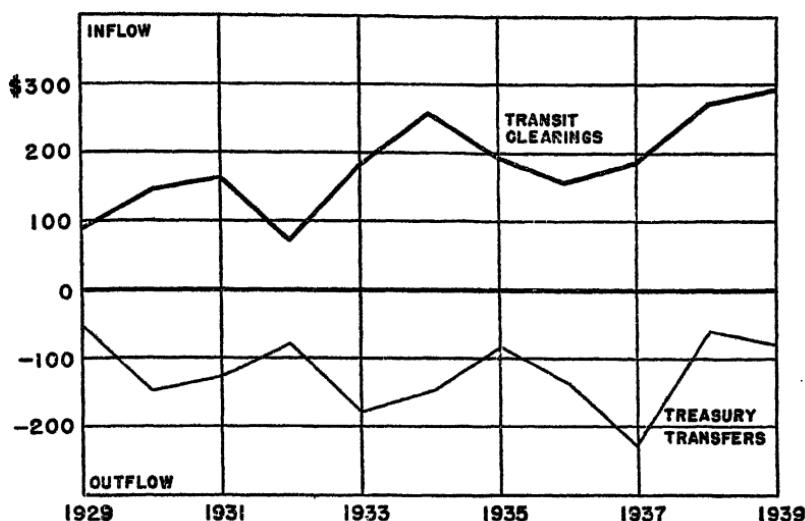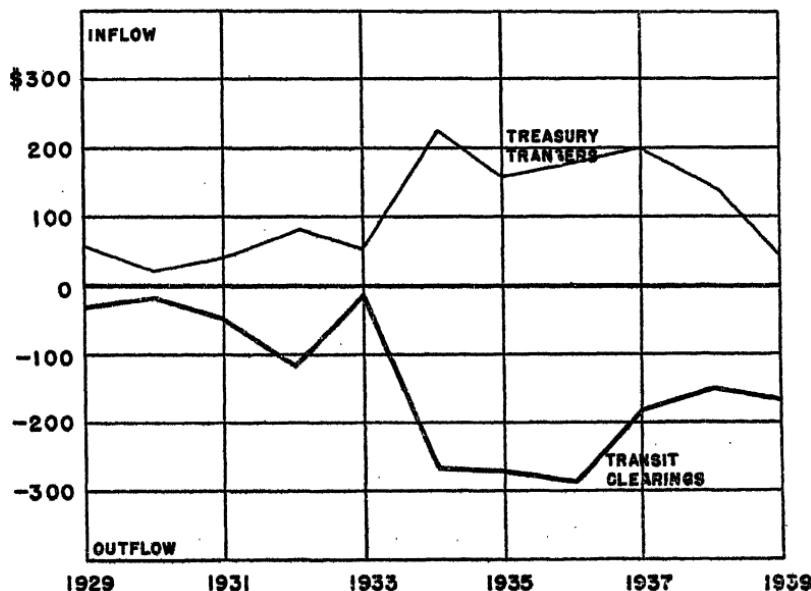
Bank. When in the course of its business the federal treasury finds that it must make payments in one region which are in excess of the funds it has on hand in that region, it will direct that part of its deposit in another Federal Reserve Bank to be transferred to the bank in the region where funds are scarce. These transfers, due entirely to government account, are termed "inter-reserve bank transfers." They are unilateral in nature, and akin to a capital movement, a gift or an indemnity in international trade. In this case, however, the transfer affects regional reserves directly, for it is effected by a movement of gold.

In considering regional equilibrium in the United States, the interesting aspect of these interdistrict clearings is the relationship, for any single district, between inter-reserve bank transfers (Treasury transfers) and transit clearings.[4] Those districts which have gained funds through the operation of the Treasury, have lost funds in transit clearings; and those regions which lost reserves on government account, have gained through transit clearings, with the result that over a period of years the net accumulation or loss of gold by any one region has been zero except during years when the gold stock of the entire country has changed.[5] The persistency of the inverse relationship is noteworthy. Since 1919, for example, the Boston district has every year lost funds through treasury transfer, gained through transit clearings. From 1919 to 1930, the New York district gained on government account, lost on clearings; while from 1931 to 1936, the New York position was reversed (see accompanying Chart). Obviously an explanation of this inverse relationship is necessary to an understanding of the conditions needed for regional equilibrium. Here we shall consider these two accounts for the period from 1929 to 1939, the era of the breakdown of the international gold standard.

### IV. THE ROLE OF TREASURY TRANSFERS

The receipts of the Treasury in any district are from two main sources: taxes (the federal income and various excise taxes), and borrowing (the sale of short- and long-term government securities such as Treasury notes, certificates, bills and bonds). The expenditures of the government take the form of payments on government

4. Since for any district the volume of Federal Reserve note clearings is insignificant as compared with transit clearings and treasury transfers, they can be ignored, and the total gain (or loss) of gold reserves by a region can be considered as the sum of the two latter.

5. See B. H. Beckhart (ed.), *The New York Money Market* (New York, 1932), II, 359. Also S. E. Harris, *Twenty Years of Federal Reserve Policy* (Cambridge, 1933), Vol. II; W. R. Burgess, *The Federal Reserve Banks and the Money Market* (New York, 1936).

contracts, service of the federal debt and federal aid payments to states and individuals.

Before the occurrence of the large relief expenditures and the deficits in the federal budget of the thirties, the largest proportion of government expenditures was made in the New York district. There, payments on many government contracts were made, and (more important in affecting the interdistrict gold movement) there, Treasury obligations were issued and redeemed and their service charges were met. According to a study of the Federal Reserve Board, redemption of Treasury obligations in New York far exceeded the allotments of the region because of active trading in these securities in the New York money market. Treasury obligations issued in other districts drifted to New York and had to be redeemed there.[6] Ordinarily about two-thirds of every Treasury issue found its way to New York by the time it matured, whereas the proportion of the federal income tax collected in New York was only one-third of the total.[7] As a result, Treasury expenditures in the New York district regularly exceeded receipts there; the difference was made up by the transfer of tax proceeds from other sections to New York. Through the twenties, therefore, and until 1931, while New York consistently gained gold through Treasury transfers, other regions lost reserves.[8]

With the transfer by the Treasury of tax proceeds from other districts to New York, the legal reserves of out-of-town banks were decreased. To replenish the latter, commercial banks in other districts withdrew balances from New York after each tax period, the withdrawn funds representing, among other things, proceeds of securities which had been sold or matured in New York and payments for merchandise exported through New York.[9] The movement of these interbank balances being accomplished through the transit clearing account of the Interdistrict Settlement Fund, the New York district consequently lost gold through transit clearings to other regions.

When purchases of government securities are made in New York for the account of other Federal Reserve Banks, payment to the New York bank is effected via the Interdistrict Settlement Fund; and when securities from the holdings of other Federal Reserve Banks mature in New York return payment is accomplished through the Fund. During this period, because of greater maturities than

6. *Federal Reserve Bulletin*, April, 1922, pp. 400–403.
7. *Federal Reserve Bank of New York, Annual Report, 1927*, p. 18.
8. With the exception of St. Louis and Kansas City, both of which gained through Treasury transfers, but very small amounts as compared to New York.
9. *Federal Reserve Bank of New York, Annual Report, 1927*.

purchases in New York, the general tendency of these Reserve bank settlements resulted in the withdrawal of funds from the second district,[1] thereby further increasing New York's loss of gold through transit clearings.

During the prosperity of the twenties, then, with a surplus in the federal budget and a decrease in the federal debt, a cumulative gain or loss of gold reserves by any region did *not* occur because of the regional activities of the federal Treasury. Those districts, whose trade and financial relationships with the rest of the world were such that their receipts exceeded their payments, were prevented from accumulating gold in the I. S. F. by a loss of Treasury balances; and those districts whose payments exceeded their receipts on commercial and financial account maintained their gold reserves as a result of Treasury spending. In brief, the activities of the Treasury increased the mobility of capital in the country; creditor regions were forced to lend; debtor districts were enabled to borrow — but with no provisions for a return flow in the future.

After 1931 and until 1937, New York's position (and that of a number of other districts) was completely reversed: New York acquired gold through its transit clearing account, but lost funds via Treasury transfers. This reversal resulted from an increase in government expenditures, largely for relief purposes, and the appearance of a government deficit financed by the issue of bonds. The banking system of the country formed the main market for these bonds, the purchases of which were made largely in New York. Meanwhile, government expenditures in other sections of the country increased, especially for public works, unemployment and agricultural relief and such, causing the Treasury receipts in New York to become greater than payments there, and thereby causing a large movement of government funds out of New York.

The volume of transit clearings and Treasury transfers of the New York district was of much smaller magnitude during 1937, 1938 and 1939 than in previous years; moreover, the direction of these clearings was reversed in each year. (See Chart, p. 398). In 1937 New York gained funds on government account, but lost funds through transit clearings. In this year, expenditures of the Treasury decreased somewhat and, moreover, were met to a larger extent out of taxes (in particular, social security taxes) than from borrowing. Relative expenditures of the Treasury in New York, therefore, increased, while relative receipts dropped, thereby necessitating a transfer of funds to the second district. In this year, the

1. *Federal Reserve Bank of New York, Annual Report, 1927.*

balances of out-of-town banks in New York decreased, resulting in an outflow of funds through transit clearings, which reflected the need to strengthen their cash position on the part of banks in districts from which the government had transferred funds.

In 1938 both government borrowings and expenditures increased (the latter mainly under agricultural, national defense and WPA projects). That the net movements into and out of New York were smaller in this year was probably due to the fact that new issues of government bonds were purchased more largely by other investors than banks, while the latter continued to invest in short-term government obligations.[2] A smaller proportion of Treasury financing was thus carried on in New York and the necessity for a compensating movement of bankers' balances was therefore smaller.

New York's loss of gold on government account (from 1932 to 1937, and again in 1938), however, was almost equalled by transfers of private funds to the city from other parts of the country. Banks in other sections were able to send funds back to New York because they were experiencing excess reserves. In fact, this return flow on transit account represented in the main deposits, by banks in other regions with New York correspondents, of part of their excess reserves.[3]

The appearance of these excess reserves in the United States at this time is to be explained in terms of the unprecedented gold inflow from abroad together with a decline in demand deposits. That these excess reserves were distributed throughout the country, however, and not concentrated in one or several districts, is to be explained in terms of Treasury transfers.

The gold imports entered this country through the port of New York, thus affecting the reserves of the second district banks first. In the words of the *Federal Reserve Bulletin* of January, 1937, "Improvement in agricultural conditions since 1933, and the increase in farm income have carried funds to agricultural regions in payment for their crops sold at higher prices. Likewise, industrial expansion in various sections of the country has resulted in a substantial shifting of funds to such sections. *Perhaps the most important factor in the geographic distribution of surplus funds of banks since 1933, however, has been the expenditure by the Treasury of funds raised by the sale of government obligations to banks.*"[4] The fact of government expenditures in excess of receipts in other sections caused the Treasury to

2. *Federal Reserve Bulletin*, September, 1939, p. 211.
3. *Federal Reserve Bank of New York, Annual Report, 1934*, p. 10.
4. P. 2 (italics added); see also *Federal Reserve Bank of New York, Annual Report, 1934*, p. 10.

transfer balances out of New York, thereby distributing gold throughout the country.

If Treasury transfers after 1931 had not assumed this new direction, it is likely that the excess gold reserves would have remained concentrated in New York, and it is also likely that agricultural districts of the United States would have suffered a cumulative loss of their existing gold reserves.

In general, it was the predominantly agricultural regions (Atlanta, St. Louis, Minneapolis, Kansas City) which lost gold through transit clearings, and to which Treasury funds were moved. The predominantly industrial regions (Cleveland, Chicago, Boston) lost Treasury funds but gained through transit clearings.

While interdistrict movements of bankers' balances would be of sufficient volume to dominate transit clearings of the New York district, this same fact is not likely to be the case for districts other than New York. Since New York is the financial center of the country, its transit clearing balance shows the effects of the aggregate of banking fund movements of all other sections. Bankers' balances are concentrated in New York. For other regions, the movement of these banking funds would represent a smaller fraction of the total of transit clearings. Hence, for districts other than New York, it is likely that the direction of transit clearings represents the direction of interregional trade, the excess of export receipts over import payments or vice versa.[5] During the thirties, then, although part of the negative transit clearing balance of agricultural sections can be explained in terms of an outward movement of banking funds, it is not likely that this short-term capital movement accounts for the complete excess of payments. Without these short-term capital movements, it is probable that the transit clearing balance of agricultural sections would also have been negative but to a somewhat smaller extent. Under these conditions without Treasury transfers to agricultural regions, gold imports would not only have remained concentrated in New York, but also agricultural sections would have lost gold reserves. This implication is warranted on two counts.

First, the receipt of gold by the United States in international transactions represented almost wholly a capital import. During the mid- and late-thirties, the current account of the nation's balance of payments was relatively in balance. Receipts exceeded payments by less than $500 million annually, while the gold flow was in terms

5. See, for example, P. C. Hartland, *Balance of Interregional Payments of New England,* doctoral thesis deposited in the Radcliffe College library, to be published shortly.

of billions.[6]  Furthermore, the capital import was in the main "hot money" fleeing from economic and political insecurity abroad, thereby being short-term in motivation and appearing in large part as a movement of banking balances.  Since, therefore, the gold movement was caused by a capital inflow, it resulted from transactions between foreign countries and financial centers of the United States, chief among which is New York.  The small excess of current receipts over payments that did exist would have occasioned some gold movement from New York to other districts in payment for their exports, but in the main gold would have remained concentrated in New York.

Second, the depression of the thirties, which caused a drastic decline of all prices and incomes, produced a far sharper deflation in agricultural industries than elsewhere.  Agricultural prices fell about 40 per cent on the average from 1929 to 1933 while prices of semi-finished and finished goods declined by 20 to 25 per cent during the same interval.[7]  During the later thirties, although the percentage of recovery was greater in agricultural than in industrial prices, the level of recovery attained by the latter was the greater.  Industrial prices, that is, more nearly reached their 1929 level than did agricultural prices, and this despite government policy directed toward agricultural relief.

Treasury transfers from New York toward other sections during this period were a reflection of the anti-depression policy of borrowing in part to relieve agricultural distress.  The expenditures of the federal government in these districts contributed to the rise of farm prices and incomes.  Treasury transfers thereby enabled these districts to maintain a certain level of consumption which otherwise would have been possible only at the expense of lost bank reserves. Especially if Treasury transfers had been continued in the pattern they assumed during the twenties, with an annual retirement of the public debt, not only would the increased gold supplies have remained concentrated in New York, but agricultural sections would have been drained of their gold reserves.  During the thirties the flow of funds from the interior to New York on government account would not long have been compensated (as it was during the twenties) by a return flow from New York in transit clearings.  The balances of out-of-town banks in New York, especially of banks in the agricultural sections of the south and west, would gradually have been withdrawn to replenish depleted cash reserves after tax periods. But it is not likely that these New York balances would have been

6. *U. S. in the World Economy*, U. S. Department of Commerce, Economic Series No. 23.
7. *Bureau of Labor Statistics Bulletin, R-1069*, p. 3.

rebuilt at the rate of the twenties through the normal course of trade, since trade was not and would not have been normal. In view of the extreme deflation of farm prices, their interregional payments on commercial account would have outrun their receipts, resulting in a cumulative loss of bank reserves.

During this period Treasury funds were transferred to agricultural districts, not only from New York, but also from other industrial regions. The fact of Treasury receipts in excess of expenditures in these other industrial regions is probably best explained by the volume of government bonds bought by them, and also by the change in the income tax structure at this time toward greater progressiveness. Since incomes in industrial areas were high in relation to those in agricultural sections, the former were on the average subject to a proportionately higher tax rate, thereby swelling further the receipts of the Treasury in these regions over their level of the twenties. The direction, therefore, of Treasury transfers reflected not only an anti-depression policy of borrowing, but also the movement toward greater equality of income.

Thus, the agricultural sections of the United States were more fortunate than the agricultural countries in South America and the Balkans during the same period. The policies of the federal government implemented by the financial activities of the Treasury caused their reserves to be replenished, making it possible for them to maintain a balance of payments equilibrium and also a certain level of consumption. Without such transfers, the loss of reserves to which they would have been subject, would have forced agricultural districts to undergo a severe deflation of purchasing power and incomes at a time when they were suffering a greater deflationary pressure than the rest of the country. Also, such a decrease in real income would have caused a migration of people from agricultural to industrial regions where real incomes were higher.

Actually, of course, both of these equilibrating forces were at work. Real incomes (consumption) in agricultural sections did decline; during the thirties there occurred a migration of the population from the country to urban districts. But in addition, the injection of purchasing power into agricultural regions by the Treasury prevented their falling as far as they might have fallen in terms of population, income and consumption.

Further, if this decline in consumption had been allowed to occur, it would have affected the level of income achieved in industrial sections of the country. Then the declining purchases of agricultural from industrial sections would have caused a lower level of business

activity in the latter with a consequent reduction of income. Under these conditions, it is likely that industrial sections, experiencing at the same time some excess reserves, would have hoarded a substantial part of their income (maintaining idle balances to a larger extent than actually they did). Thus, the anti-deflationary spending of the Treasury in mobilizing and distributing bank reserves throughout the country, raised total income and consumption above what it would otherwise have been.

## V. SUMMARY

To conclude, the successful operation of the interregional gold standard within the United States during the period which saw the collapse of the gold standard internationally, seems to be best explained in terms of factor mobility: the fact that people moved freely between agricultural and urban localities in this country and especially the fact that capital was given greater mobility by the activities of the federal government. The mechanisms of adjustment available to any economic region for meeting a balance-of-payments disequilibrium (given fixed exchange rates) lie in an income change, a population change, or a capital movement. The more the adjustment is made via factor movements, the smaller will be the necessary adjustment of income. In view of the different degrees and types of industrialization which characterize the trading regions of the world, homogeneous economic policies applied to all regions without discrimination are more likely than not to be upsetting to a balance of payments equilibrium. It is interesting to note that Alfred Marshall, who, like Aristotle, had a word to say on almost every conceivable topic, described the difference between international and interregional trade in terms of differences in factor mobility and the fact that "every part of a country is called upon to contribute to the expenses of its central government."[8]

PENELOPE HARTLAND.

BROWN UNIVERSITY

8. *Money, Credit, and Commerce*, p. 155.

# THE COMMONWEALTH COURT OF CONCILIATION
# AND ARBITRATION: A BRIEF SURVEY

## SUMMARY

## I. INTRODUCTORY AND GENERAL

The Commonwealth Court of Conciliation and Arbitration of
Australia came into the world in 1905 unwanted by an important
and influential section of industrialists — by many employers who
regarded its appearance as an intrusion and an impertinence, as well
as a direct threat to their material well-being.   These feelings of
hostility or vexation did not quickly dissipate.   Entrepreneurs of
that mind boycotted or avoided the Court to the greatest extent
practicable, endeavouring to dispose of all matters between themselves
and their employees by private negotiation.   If brought before the
Court, they questioned its jurisdiction and attempted to obstruct or
delay the hearing at every available opportunity.   Should the decision
be not to their liking they sought to defeat its enforcement, and
render an award a nullity, by application to the High Court for the
issue of a writ of prohibition, alleging that the Arbitration Court had
exceeded its powers and was usurping jurisdiction, or that the legisla-
tion supporting it transgressed the constitutional limitations of the
Federal Parliament.[1]   Their opposition was mitigated only when it
was evident that the Arbitration Court meditated no dark designs
on managerial rights, and that those rights would not be modified
apart from considerations of justice to the worker, and of the general
well-being.   A factor in the changing view, no doubt, was the realiza-
tion that the Court was acting on principles, the consistency in the
application of which boded promisingly for fair and equal treatment
for both parties in industry, a more amicable feeling in employer-
employee relationships, and a protection for decent employers
against the activities of less scrupulous business rivals who, otherwise

---

1. It was not till the passing of legislation in 1914 that a justice of the High
Court was enabled to give a decision on a point of law or jurisdiction *before* the
conclusion of the hearing of the dispute or proceeding before the Arbitration Court.

paying at less than award rates, might undersell them in a competitive market.

Nor was there any exhibition of pleasure from other jurisdictions at the arrival of an authority whose presence signified the possibility of profound change in the basis on which agreements between employer and employee rested. Not that they entertained the hard and unnatural sentiments traditionally attributed to the elder sisters of Cinderella. But the older lawyers were suspicious, and inclined to look somewhat askance at the newcomer as the offspring of a heretical conception. For, all said and done, the industrial jurisdiction in Australia, with its impositions of terms and conditions of employment, does cut right across the common law notions and doctrines of contractual freedom as far as the relation of master and servant is concerned. The general public, too, was little moved by the event, although the new jurisdiction represented a portentous step for the safeguarding of the communal interest as well as the interests of those occupied in industrial pursuits — a recognition that peace in industry and the adjustment of the terms and conditions of employment are of supreme importance to the people as a whole, and not matters of mere local or private concern.

The nearest approach to an indication of satisfaction emanated from the trade unions. They had been disastrously defeated in recent years in their contests with employers, and it was primarily for the protection of the workers, and the recovery of their fundamental rights, that the Commonwealth Conciliation and Arbitration Act was passed in 1904, and the Court thereby brought into existence. But the welcome that was accorded by the unions was, perhaps, lacking in warmth and enthusiasm. Union leaders were conscious of the true implication of compulsory arbitration in industrial disputes — the logical irreconcilability of authoritative regulation with direct action or any other method of self-redress — and they were uncertain and uneasy at a prospect where the strike weapon was not available. Nevertheless, the unions showed a desire to become organizations registered under the Act in order to participate in the advantages of control. The legal requirements were not exacting. Certain conditions are particularized in the Act and, in the regulations promulgated under the Act, and on the fulfilment of these, any association of not less than one hundred employees in (or in connection with) an industry may be admitted to registration as an organization. A union that satisfies the prescribed conditions has every expectation of success in its application for registration, unless already there is registered another organization representative of workers in the same

field — another organization to which members of the applicant union might conveniently belong. The same considerations apply to any association of employers in (or in connection with) any industry, who have in the aggregate, or any employer who has, throughout the six months next preceding the application for registration, employed on an average taken per month, not less than one hundred employees in that industry. The conditions with which compliance is demanded as a pre-requisite to registration are such as, in general, could be met by any well-ordered body of employers or employees whose rules are framed in the interests of majority control by its members. The exercise of the power of the Registrar of the Court to refuse registration where there is in existence another registered organization, even if the conditions prescribed in the Act and under the regulations have been observed, is justified by the belief that industrial peace may be prejudiced by the presence of more than one organization as spokesmen for the same interests in an industry.

In effect, by registration under the Act, an organization becomes a corporation with power, inter alia, to purchase real or personal property, or to take it on lease, let, sell or exchange it. It enjoys, in practice, a monopoly of the privilege of organizing, and representing before the Court, employees engaged in the occupations or industries covered by its constitution. To it accrues the privilege to obtain awards in the interests of its members, and to take action in enforcement of their rights derived from such awards.

Registration as an organization under the Act is purely voluntary. It is not generally understood that there is no compulsion in law upon any union or association to make application for it. And just as registration springs from the free choice of a union or association, so a continuance of that registration depends, in effect, upon the will of the organization to retain that standing. Should an organization change its mind and prefer to break loose from federal industrial control, its unwillingness to maintain the connection will, in practice, not go long unrecognized. Refusal to obey the orders of the Court may result in the Court ordering the cancellation of its registration. Theoretically, such a cancellation is not necessarily in the nature of a penalty or sanction; it is an acceptance of the fact that the organization has repudiated the object and effect of registration, and desires to withdraw. De-registration involves the loss to the organization (and to its members) of title to the benefits of any award made under the Act, deprivation of status as a corporation under the Act and the sacrifice, as well, of other privileges accruing under the Act such as the right

to sue for and recover, in its own name, fines, fees, dues or levies payable to it by members under its rules.

Orders for cancellation of registration have been few in number, and they have been made only with reluctance. Even when the order is made, the Court is disposed to let it "lie in its registry" for some little time — that is, defer its operation — in the hope that, at the eleventh hour, the organization will see the foolishness and short-sightedness of its actions and come to a different decision. Where the hope does not materialize and the order takes effect, the Court can be confident of a change eventually in the attitude of the union or association. In fact in every case where registration has been cancelled, the union or association has soon after sought re-registration. The latest example is the Building Workers Industrial Union of Australia, which, under the Court's order, was de-registered as an organization of employees as from August 27 of last year, and which lodged an application for re-registration on October 27. It is the Court's practice to accede to these requests where assurances are forthcoming that orders and awards will be observed in the future. The Court, shrewd in its experience, is aware that human aggregates have their bouts of insubordination and restiveness with the things that be. It knows that, at times, men will be persuaded that they are longing for "freedom," and that they are capable of managing their own affairs, and fighting their own battles, without the aid of authority. It knows too, from the store of its wisdom, that these periods quickly pass when the distant fields are discovered to be not so green as they appeared, and the benefits and privileges of registration are seen in another light.

The action of the Building Workers Industrial Union of Australia in seeking re-registration is pregnant, perhaps, with more than usual significance. At present, with labor in short supply and the Government determined to maintain conditions of full employment, the services market is essentially one for sellers. There has been a complete change in the situation of earlier days; it is now the employers who would seem to be in need of the protection that formerly was sought by their employees. In such circumstances the practical value to a union of registration under the Act, and submission to legal regulation, might seem to have been considerably diminished — or even to have disappeared. Conclusions such as these might have had their appeal to those responsible for the direction of the affairs of the Building Workers Industrial Union of Australia, and the determination of the union to take its own course. Its reversal of policy, and the decision again to march with the industrial tribunals, serve

to emphasize that, in times favorable to employees equally with those suiting employers, a system of authoritative industrial regulation can perform its designated function.

Indeed the volume of latent support available to the federal arbitration system is open to grave miscalculation. A lukewarm attitude towards industrial control is, with some organizations, merely a cloak or a bargaining point. In their view it is not good tactics to allow their true feelings on industrial regulation to appear on their sleeves. Petulance aside, this is the explanation of much of the disgust that, at times, has been expressed with the content of awards, and of the protestations of a preference for direct action. One has but to recall the *denouement* of the Maritime Industries Bill that was introduced into Parliament in 1929 for the repeal of the Commonwealth Conciliation and Arbitration Act 1904–1928 (with the consequential abolition of the Court) and the vacation by the Commonwealth of its industrial jurisdiction.[2] Nothing has happened to show that a party sponsoring such a measure will not, as in 1929, succeed in bringing down upon itself a political landslide, and the righteous indignation of an angry people of whom the most outraged will be counted from among trade unionists themselves.

## II. Arbitration and Conciliation

It has been indicated that the reception accorded to the Court at birth was not unmixed. The duties, too, cast upon it by the Commonwealth Conciliation and Arbitration Act were difficult in the extreme. In their fulfilment its president had before him the colossal task of ascertaining standards appropriate to Australia, and of molding and consolidating them into a jurisprudence. The problems involved were complex and multifarious; they were social and economic as well as legal and constitutional in character, and their solution demanded at the very least a working knowledge of industrial techniques. With New Zealand (since the enactment of the Industrial Conciliation and Arbitration Act in 1894), and New South Wales (with the passing of the Industrial Arbitration Act in 1901), practically the only countries with a regime of compulsory arbitration, the guidance to which the newly established Australian Court could look was limited, the more so because in those countries, there was little or no experience of the jurisdictional dilemmas which the division of an industrial power between two authorities (federal and state) can occasion.

2. See the author's *Towards Industrial Peace in Australia* (Melbourne University Press, 1937), pp. 55, 56.

The Court was instructed by the Act to be expeditious in the making of its decisions, but union officials in the role of advocates appeared before it largely, and their lack of a legal training proved a serious impediment to the conduct of a case with directness and conciseness. As an arbitrator the Court was probably not competent constitutionally to refuse to hear argument however prolonged, or reject evidence however protracted, provided the argument or the evidence were not irrelevant to the matter in dispute.[3] Its jurisdiction, too, was restricted to industrial disputes extending beyond the limits of any one state, and, in the discharge of its responsibility for the prevention and settlement of these disputes, the only methods permissible were conciliation and arbitration. It took the High Court many years to decide, with something approaching finality and only after a tremendous volume of bewildering litigation, the frontiers of this jurisdiction and the extent of the Court's powers. However it was demonstrably established in 1910, in *Whybrow's Case*,[4] that the Constitution forbade the investment of legislative authority in the Court so that, in spite of the prima facie liberality of the Act, the Court was not competent, having made an award in settlement of an interstate industrial dispute between the parties, to proceed to declare that award binding as a common rule for the industry within which the dispute arose.[5] Compulsory arbitration, in its true character, connotes the bringing before the arbitrator by compulsion of the parties to the dispute, a hearing of them as to the matters in dispute, an award in determination of the disputed matters, and the binding of the parties by that award. The consequence of the decision in *Whybrow's Case* meant that service of process had to be effected on all respondents whom it was desired to make parties to awards, though this might involve expense, irritation and delay in the obtaining of that award.[6]

But that was not all that was to be understood from *Whybrow's Case*. Its doctrine suggested that non-unionists were beyond the

3. Compare *Australian Railways Union v. Victorian Railways Commissioners and Others*, 44 Commonwealth Law Reports, 1930, p. 319.

4. 11 C. L. R., p. 311.

5. Section 38(f) of the Act purported to confer this power. It was declared *ultra vires* and invalid.

6. There is a commerce power available to the legislature under the Constitution the same, in substance, as that in the Constitution of the United States, with the method of exercise *not limited to conciliation and arbitration*. Australian courts, however, have not followed American courts in attaching to this power a broad interpretation. The regulation of industry at the hands of Australian federal tribunals rests strongly upon the exercise of the constitutional industrial power.

originally comprised the High Court, the jurisdiction of the Arbitration Court was, in effect, confined to industries in which there was interstate competition. Had this conception of the Court's jurisdiction not been rejected, the Court would have been precluded from regulating the terms and conditions of employment in important industries such as the building trades, municipal enterprises and tramway undertakings. In the *Engineers Case* (1920),[9] the *State Railways Case* (1906)[1] was overruled, the Court deciding that an industrial instrumentality of a state could be made party to a federal award in settlement of an interstate industrial dispute in which it was involved, and bound as to the terms and conditions on which its employees worked.[2] The decision in the *Engineers Case* was of increasing practical significance because of the tendency of the states to extend their operations in the field of private enterprise, or expand their activities into areas not strictly "regal." In *Cowburn's Case* (1926) the Court, called upon to interpret section 109 of the constitution, held that, contemporaneously with an award made by the Arbitration Court in settlement of a matter in dispute — whether positively by allowing the claim, or negatively by refusing it partly or in toto — any state law, award or determination dealing with the same matter could not operate.[3] A field once entered by the federal industrial tribunal was deemed occupied by it, and fully closed to the effects of state industrial decisions, as long as the federal award remained unterminated. This decision, together with the desire of Commonwealth and state industrial tribunals for co-operation and comity, went far to remove the embarrassments and inconveniences that had sprung from the overlapping and duplication of Commonwealth and state awards. The situation, in this respect, was further relieved by legislation in the states of New South Wales and Victoria providing that the basic wages under awards or determinations of those states — the most populous and the most highly industrialized in the Commonwealth — should be coincident with the Federal

9. 28 C. L. R., p. 129.

1. 4 C. L. R., p. 488.

2. In the *State Railway Servants Case* the High Court was strongly influenced by American decisions — notably *McCulloch v. The State of Maryland* (4 Wheat., p. 316) and *The Collector v. Day* (11 Wall., p. 113). In the *Engineers Case* the doctrine of the immunity of State industrial instrumentalities, enunciated by Chief Justice Marshall in *McCulloch v. The State of Maryland*, was discarded as inapplicable to the interpretation of the Australian Constitution.

3. 37 C. L. R., p. 466. Section 109 provides that where there is an inconsistency between the law of a state and a law of the Commonwealth, the latter shall prevail.

Court's basic wage; indeed, the Victorian legislation instructed wages boards to incorporate other provisions of federal awards as well in their determinations, as far as permissible under their powers.

Amendments to federal legislation in 1926, which provided a life tenure for judges, enabled the Court to exercise the judicial power that Parliament had previously, in vain, attempted to confer upon it. The constitution provides that the federal judicial power is exerciseable only by those appointed for life, and for that reason the High Court had held in *Alexander's Case* (1918) that, in spite of what the legislation had to say, the Court had not been invested with any such authority.[4] The president and deputy presidents of the Court had not been appointed for life. After 1926, however, constitutional objection could no longer be taken to the exercise by the Court of the power to impose penalties, and interpret awards, provided for in the Act — a power of considerable practical convenience in the operation of any scheme of industrial regulation.[5]

Fortified by the liberality of High Court decisions, strengthened by amendments of the Act aimed at an improvement in the regulative machinery, its influence extended by the accommodating provisions of state legislation, and fortunate in the judges who directed its destinies, the Federal Court developed into the outstanding industrial tribunal of the country. It had become organic in Australian social and industrial life — part of the order of things in Australia. If secondary as well as direct permeations are taken into account, there could be no doubt that the Court had become the most determinative influence in the working lives of the great majority of Australian employees.

### IV. THE GROWTH OF THE COURT'S PRESTIGE

It is generally admitted that, by and large, the Court handled, with skill, the situations and problems with which it was confronted during the First World War. These arose principally out of the fall in the workers' standards of living attributable primarily to currency inflation. The Court, too, emerged from the ordeal and the testing fires of the years of very severe depression (1929–1933) with a prestige and a reputation in many respects enhanced, albeit its wage-fixing policy in the crisis was in the light of subsequent events shown to have been probably mistaken. Generalizing, it can be said that the

4. 25 C. L. R., p. 434.
5. See the author's *Industrial Regulation in Australia* (Melbourne University Press, 1947), pp. 54, 55.

Court, by its work over the period from its inception to the outbreak of war in 1939, had established itself strongly in the public confidence and captured its imagination. There was an appreciation of the part that it played as a custodian of the communal well-being, and as an organ of distributive justice among industrialists. While no one could claim that it had fulfilled all the hopes centred in it by some of the early federal parliamentarians, and while there were occasional, almost inexplicable failures at its hand, the Court's reactions to new ideas, and the tremendous influence that it exerted on the national economic life, were recognized as, on the whole, in the direction of a general and very considerable betterment. Through its agency the workers had, in the main, been protected against exploitation; living standards had been raised without demonstrable injury to employment; industrial stoppages reduced in number and industrial warfare mitigated; the expansion of secondary industry encouraged and assisted; a more equitable basis of remuneration established between the various occupations and industries by way of wages to the employee and profits to the employer; a powerful stimulus applied to collective bargaining through registered organizations and an important contribution made to the institution of a balanced and stable national economy. It would have been futile to assert that Australia had not gained from her experience of compulsory conciliation and arbitration.

Criticism sprang chiefly from the unions. In so far as this criticism was addressed to techniques, much of it undoubtedly was soundly based — thus the relative insufficiency of the part played by conciliation in the settlement of industrial disputes, the necessity of investing the Court with "cognizance" of a dispute as a condition to the exercise of its arbitral power,[6] the necessary delays in the actual hearing of cases and the length of the hearings. In so far as it was directed to the decisions of the Court, it was, however, often ill-conceived and emotional. At times it travelled beyond the concession that is allowable to irrationalism and impatience; for example, some union leaders alleged that the Court had preconceived notions on the desirability of a forty hours standard working week, rendering it a matter of little hope to submit the question for arbitration.

6. Four channels were specified by the Act through which "cognizance" could be obtained. Usually it was given by the following formal steps — service of a log of claims on the other party to the dispute, refusal of the log, the holding of a compulsory conference before a judge or a conciliation commissioner, the failure to reach agreement at the conference, and the reference of the dispute by the judge or commissioner into court.

## V. The Second World War

The world catastrophe of 1939 brought with it, for the Court, a responsibility that probably could not, in the circumstances, have been discharged with satisfaction by any industrial tribunal. By virtue of certain legal dispositions its powers were certainly reinforced on the whole for the occasion, and other authorities were introduced to take some of the responsibility from its shoulders. Under the National Security (Industrial Peace) Regulations, the support for which lay in the exercise of the legislative defense power permitted by the constitution to the federal Parliament, the Court was invested with the common rule power held to have been unavailable to it under the industrial power.[7] By the same regulations it was enabled to make awards in settlement of intra-state industrial disputes. Jurisdiction to make an award could also be conferred upon it under these regulations by the Minister of Labor and National Service when any industrial matter had led, or was likely to lead, to industrial unrest, notwithstanding that an industrial dispute affecting the matter did not exist. As a matter of fact much important business that engaged the attention of the Court during the war years was derived from ministerial reference. And then, under other regulations issued in pursuance of the National Security Act, or by virtue of legislation in direct exercise of the constitutional defense power, administrative authorities were appointed to assist in the task of industrial regulation — thus the two successive Women's Employment Boards, the Stevedoring Industry Commission, the Maritime Industry Commission, and the special authorities established in relation to the coal mining industry. The Women's Employment Boards were created partly in pursuance of the war-time policy of maximum production, in an attempt to remove complaints against delays in the hearing of claims of women workers, large numbers of whom had been imported into industry to cope with existing labor shortages. Their jurisdiction extended to defined classes of female workers, notably those taking the place of male workers required for service in the country's defence. The constitutional justification for them was alleged to be the need, as incidental to the defense power, of securing the return to their avocations after the war of men serving in the forces. The Stevedoring Industry Commission was invested with jurisdiction as to the waterfront, in the hope that, with an administrative authority devoting its whole time to the demands of

7. The High Court allowed a very broad interpretation to be applied to the defense power in the emergency conditions raised by the war.

the workers in that area, there would be a quicker "turn round" of the ships, so vital to the prosecution of the war. Conditions in which sea-going employees operated (the traffic in danger zones, etc.) were seen to have changed from those in which their duties were performed in normal times, and it was believed that workers would become more contented, and efficient, with the presence of an administrative authority (the Maritime Industry Commission) that could apply, forthwith, principles conforming with the new circumstances. The coal mining authorities were established in a new attempt to terminate the old feuds between colliery proprietors and their employees, and to augment the output of a commodity upon which, no less than upon the labor supply, the measure and tempo of the Australian war effort depended.

The relief that these emergency "defense" authorities brought to the Court was, however, by no means unqualified, in that the delimitations of the new jurisdictions were not always clear, the authorities worked independently of the Court, and, of their own choice or by the terms under which they were appointed, they applied principles that, after allowing for difference of circumstances, were not always identifiable with those followed by the Court. Much discontent was caused when workers governed by Court awards observed that, without apparent justification, more favorable terms and conditions of employment were granted to others for services recognized as similar to, or comparable with, those that they themselves were performing.[8]

Still other regulations under the National Security Act withdrew from the Court its competence, in important respects, to move wage rates and alter conditions of employment. Such regulations were those pegging wage rates unless an anomaly or change of circumstance could be shown to exist (in the interest of the national policy of all-round stability), fixing the maximum hours to be worked (in the interest of the workers' health, and to reduce absenteeism in the factories), providing for dilution of labor in industry (to help solve the problem of a scarcity in skilled workers), and those curbing the right of employers to dismiss employees and of employees to leave their employment (to economize in labor, and to apply it to the best uses).

In the new industrial conditions caused by the war, the pressure

8. For a further description of these emergency bodies, see the author's *Solving Labour Problems in Australia* (Melbourne University Press, 1941), pp. 153 et seq., and his *Wartime Labour Developments in Australia* (Melbourne University Press, 1943), ch. V.

lay overwhelmingly on labor, whose services were required insatiably by both the combatant forces and the factories. With the strain upon labor so intense, more than ordinary tact and diplomacy, as well as wisdom and strength, were needed in the efficient discharge of the Court's function. Mistakes the Court did make — and serious and substantial ones at that. The policy of uniformity and consistency that had served it so well in the past in the making of awards — the only policy with any prospect of success for an arbitral tribunal to embrace — was too often departed from, and surrendered, at the dictates of opportunism. But there was ever present, of course, the paramount consideration of the need for victory over the enemy — for which even the price of industrial appeasement would have to be paid, if necessary.

### VI. THE POST-WAR DEVELOPMENT

Labor was in charge of the administration at the close of the war with full political opportunity, within the constitutional framework, to reshape the Commonwealth industrial regulative machinery according to its own ideas, and in the light of experience. There was a strong feeling among the workers for a continuation of the exceptional treatment of the waterside, sea-going and coal mining industries. Consequently, in disregard of the dangers inherent in an arrangement that breaks down the principle of a single unified control throughout all industry, it decided for a policy of special authorities for the regulation of these industries in normal times. Legislation enacted in 1946 purporting to be under the ordinary constitutional power has partly effected this purpose in the coal mining industry, and, in 1947, fully as regards the waterfront, while, with respect to the sea-going occupations, a measure will probably be not long delayed.[9] The constitutional validity of such legislation, however, is still open to attack in the High Court.

To cover the generality of workers, an amending Conciliation and Arbitration Act was passed in 1947, aimed mainly at removing long-standing criticisms by the unions, some of which have been alluded to in this article. Under it, the requirement of "cognizance" as a pre-requisite to the exercise of jurisdiction was abrogated. Provision was made for the establishment of a research bureau, the results of whose investigations should, by reducing the quantum of evidence that otherwise might be submitted by the parties, sub-

9. The second of the Women's Employment Boards had been disbanded in 1949.

stantially shorten the time needed for hearings.[1]  To reduce delays in the beginnings of hearings and to allow of a greater prominence for the method of conciliation, reliance was placed on an increase in the number of "conciliation commissioners" and the establishment of a new relationship between these commissioners and the Court. A commissioner was enjoined, should it appear to him that an industrial dispute had occurred, or was likely to occur, to ascertain immediately the parties to the dispute, and the matters that were in issue. He was bound then to take such steps, by way of conciliation or arbitration, as he deemed proper for the prompt prevention or settlement of the dispute.  The same obligations rested on him should he become aware of an industrial situation likely to cause an industrial dispute, or if there were reported to him an industrial dispute (actual or threatened), or an industrial situation calculated to lead to an industrial dispute.  In proceedings before the commissioners no party was to be represented by counsel, solicitor or paid agent; as regards the Court, if the proceeding were industrial, counsel, solicitor or paid agent required the leave of the Court and the consent of all parties as a condition to appearance, but if the proceeding were judicial, counsel or solicitor could appear as of right.  The administration believed that the achievement of the best interests of the workers, and the smooth operation of the industrial regulative system depended primarily on the changes effected under the Act, relative to the powers and jurisdiction of the Court and the office of conciliation commissioner respectively.

The office of conciliation commissioner, dating from legislation of the later 'twenties, was of comparatively minor significance until the year 1940.  Not more than three commissioners could hold office simultaneously, but the National Security (Industrial Peace) Regulations, issued in 1940, permitted the appointment of an unspecified number.  The office grew in usefulness during the war, although continuing purely as a subordinate and an auxiliary to the function of the Court — an attachment to it.  The practice was for a judge to depute a commissioner to inquire into a dispute, actual or threatened, and on his report being received, to decide whether himself to adjudicate or to instruct the commissioner to settle the trouble by conciliation or arbitration.  Under the Act of 1947, however, the

1. The functions of the bureau are specified by the legislation as (a) to collect and compile, in accordance with the directions of the Chief Judge of the Court, information which may be of assistance to the Court and to conciliation commissioners in the exercise of their powers and functions under the Act; (b) to keep information so collected and compiled up to date: and (c) to carry out research in respect of such matters as the Chief Judge indicates.

office was completely reorganized and elevated to a footing that renders it essentially independent, in its operation, of the Court, and in possession of a goodly portion of the jurisdiction that once belonged to the Court. For the Court was deprived of arbitral jurisdiction in all matters except the alteration of — (a) the standard hours of work in an industry: (b) the basic wage or the principles upon which it is computed; (c) the period which shall be granted as annual leave without pay; and (d) the minimum rate of remuneration for adult females in an industry.[2] These issues are reserved to the exclusive jurisdiction of the Court. They were singled out by the legislature as matters in regard to which, by reason of their complexity and overriding importance, there should be general uniformity throughout industry. Apart from these issues and the regulative powers of the special authorities, conciliation commissioners (at the moment fifteen in number) have sole conciliative and arbitral power over practically all industrial matters within the Commonwealth jurisdiction. No qualifications are stated in the Act for appointment to the post of conciliation commissioner, as in the case of the Court (where a legal attainment is required), and their number is legally capable of indefinite increase.

It is the right of the Chief Judge to assign a conciliation commissioner to a particular industry or group of industries, or to a particular industrial dispute. This right has been exercised, and each commissioner allocated to a specified industry or a group of industries. Had the Chief Judge not made the nominations it would have devolved upon the Chief Conciliation Commissioner, as a duty, to organize and distribute the work of all his colleagues. There is no requirement of a legal nature on the commissioners to collaborate in their conciliative or arbitral efforts in the prevention and settlement of industrial disputes; no compulsion of any kind can be applied to ensure a uniformity in their procedure, methods or standards. The discretion of a conciliation commissioner is practically unfettered.

The duties of the Court, and of the commissioners, are intended to be mutually exclusive and carefully delimited. Certainly, under the Act, an obligation is cast upon the Chief Judge to summon a conference of the commissioners, not less frequently than once in every four months, for the purpose of discussing matters relating to the operation of the Act and, in particular, means to ensure expedition

2. Further legislation (in 1948) enabled the Court to *determine,* as well as alter, the minimum rate of remuneration for adult females in an industry. It happened that while standards covering (a), (b) and (c) had been fixed by the Court, there was no corresponding prescription for (d).

in the settlement of industrial disputes. But no binding direction can be given by him, and a commissioner is at liberty to accept or reject any advice that is offered. A commissioner's independence of the Court is qualified only to the extent to which he elects to exercise his right, under the amending Act, to submit certain questions of law or of jurisdiction for its ruling. In that event he is compelled to frame his award in conformity with the ruling, or, if the award be framed before the ruling has been made, to modify the award accordingly. No compulsion is laid on him to seek this ruling: he is free himself to decide any law points arising in a matter being dealt with by him even though, because of the ban on the appearance of counsel, solicitor or paid agent before him, elucidatory assistance is not available. As a matter of fact, three only of the commissioners now officiating in relation to the general body of industry are legally qualified men. In no circumstance does an appeal lie from a commissioner to the Court — there is no provision of any kind for supervision of his operations or revision of his work. The old inferiority of status has been removed. The commissioners are no longer short-term office-holders; the five-year tenure under the pre-1947 Act is changed to appointment until the age of sixty-five is attained.

A main feature of the legislation of 1947 was, in short, the conversion of the conciliation commissioners into a distinct part of the Commonwealth arbitral machinery in dissociation from the Court. Conferences and hearings before a commissioner are not proceedings of the Court. His decisions are something apart from a judgment of the Court. For those decisions and for the work of the commissioners generally, no responsibility of any kind resides in the Court. The Court is without power to exercise, in a direct manner, a co-ordinating influence upon their work and decisions. It was visualized by the legislators that the commissioners would discharge their functions and duties, and exercise their powers, within a scheme and method completely differentiated from the operations of the Court. And, even apart from the Court, there is no co-ordinating influence imposed upon them.

The Court's judicial power was not diminished by the alterations effected in 1947; in fact, its purely legal authority was added to. But, whereas under the old law a judge, except as to alterations in standard hours and in the basic wage (or the principles on which it was computed), exercised all the powers of the Court (both arbitral and judicial), now, only with respect to a prescribed matter of practice or procedure is he, sitting alone, competent to exercise the jurisdiction of the Court. As regards all other matters, the jurisdiction of the

Court is exercisable only when it is constituted of not less than three of its judges. This is in contrast, it will be observed, with the provision permitting the powers and functions allowable to the commissioners to be exercised by a single commissioner, and it certainly runs counter to the accepted notions of what should be entrusted to a single judge. No judicial power is reposed in the commissioners. Parliament would have been compelled to admit them to a life appointment if it had desired that they should exercise this power.[3]

### VII. The Report of Chief Judge Drake-Brockman

It would be a mistake to assess the changes of 1947 as representing the last word on the subject of authoritative industrial control in Australia. On the contrary, it seems clear that some of these amendments were misconceived, and will have to be revoked or modified. There is reason to believe that, although many proceedings dealt with by conciliation commissioners have been "speeded up," those before the Court have been impeded by the statutory insistence on the presence of at least three judges in the hearing of practically all matters within its jurisdiction. Formal arbitration, too, has not been relegated to the distant background visualized by the legislators of 1947 — it is not reasonable to expect this development as long as the sequential connection with conciliation remains unqualified. In the first of the annual reports submitted by the Chief Judge in conformity with his obligation under the Act, interesting comment is directed inter alia to the lack of uniformity in the activities of the commissioners (described as "disturbing"), to the effects of the enforced absence of legally qualified men and other skilled persons from proceedings before the commissioners, to the condition of employer-employee relationships in industry and the matter of observance of awards, and to the results of the vesting of the Court's jurisdiction, for all intents and purposes, in more than one judge.[4]

Referring to the failure of the commissioners to achieve a measure of uniformity in the approach to their problems, the Chief Judge says that some of them "appear to be more influenced with the desirability of preserving a continuity of principles long established and found to be just and equitable than others upon whom the previous decisions of the arbitration authority have exerted less influence." He adds: "It may be necessary, and indeed it has been urged by conciliation commissioners in conference, that a measure of co-ordination, to be

3. See *Alexander's Case, ubi supra.*
4. The report covers approximately the first year of the operation of the Act as amended in 1947.

provided by some form of appeal, at least in matters of general importance, should be included in the provisions of the Act.[5] There is as great a practical necessity for uniformity in such matters as, for instance, the principles upon which marginal rates, shift-work rates, overtime and special-time rates, the spread of hours of work and many other conditions of employment, should be determined, as there is in such matters as those mentioned in section 25.[6] There have been instances where a commissioner has taken an individualistic line in respect of such matters with the consequent risk of dissatisfaction and dispute in other industries than that with which he has been dealing. Certainly a measure of co-ordination in respect of all matters of principle should be achieved; but this appears to be impossible unless there be established some co-ordinating authority. I cannot emphasize too strongly the desirability of effecting an authoritative co-ordination. Within co-ordinated principles the discretion of the arbitrator can be maintained."

The Chief Judge concedes that the commissioners "appear generally to have brought to their task a desire to expedite the bringing on of matters properly within their assignments, to hear the parties promptly and be as prompt in issuing directions and giving their decisions after the hearing." But, he says, where an arbitral hearing and determination have eventuated in consequence of a failure of the conciliative process, "necessarily very much the same time has been taken as would have elapsed in like proceedings before the Court as previously constituted" (i.e., of one judge). It is in this regard that the remarks of the Chief Judge, anent the ban on the appearance before conciliation commissioners of members of the legal profession and paid agents of special skill in industrial matters, are made. Their exclusion, he believes, has increased a commissioner's difficulty in his "not inconsiderable task of ascertaining the facts and obtaining the information he must have." He continues: "Some commissioners have by one subterfuge or another endeavoured to evade this prohibition in the Act. It seems to me that commissioners generally would welcome the right to allow parties to industrial disputes to be represented by skilled, experienced and qualified

5. The Act, before the amendment of 1947, provided for an appeal to the Court, constituted of the Chief Judge and not less than two other judges, against any provision in the order or award of a conciliation commissioner affecting wages, hours, or any other condition of employment which, in the opinion of the Court, was likely to affect the public interest.

6. This is the section enumerating the four matters deemed of major importance and retained to the conciliative and arbitral jurisdiction of the Court.

representatives including solicitors or counsel. I would also say that a large number of organizations would ... The generally held view of commissioners is that these provisions deprive them of the skilled assistance to which they should be entitled and which they most urgently need. I am of the same opinion. I feel also that the general effect is that the smaller unions of employees and employers are the more seriously disadvantaged ... It is thought that much good would come from the point of view both of expedition and satisfaction if power were given both to the commissioners and to the Court to allow representation by skilled and qualified representatives, whether so called industrial advocates or solicitors or counsel. Experience has shown that the function of conciliation is more frequently aided than impeded by the presence at the conciliation table of such representatives.[7]

The Chief Judge confessed his inability to comment on the extent to which, if at all, greater good will in industry has been promoted as the result of the arrangement introduced in 1947. Nor, he adds, does he possess evidence as to any alteration in the manner in which the operation of awards has been encouraged. But he does say: "The power to enforce awards and orders against employees still remains ineffectual. There have been too many instances of the refusal to accept the determinations of conciliation commissioners made in settlement of claims."

The Chief Judge is unhesitatingly critical of the provision requiring the jurisdiction of the Court to be exercised by a bench of not less than three judges, irrespective of the importance of the case to be heard. In his opinion it does not make for expedition. He thinks it "perhaps desirable" that the Court, in sitting as an appellate tribunal, should be consituted of at least three judges, but in cases that depend on the hearing of evidence, and the determination of facts, he believes that a multiple tribunal is "generally speaking inappropriate and involves delay." Jurisdiction in relation to prosecutions under the Act, or to applications for penalties for breaches of the Act or of awards, or to interpretation of awards should, he considers, be invested in a single judge with a competence in him to co-opt his colleagues to sit with him should such a step appear to him desirable.

7. The smaller organizations cannot afford to employ a person occupied solely on industrial advocacy, and, in any case, they do not have enough work to enable an officer to acquire a sufficiency of knowledge and experience in that direction.

## VIII. Concluding Remarks

The main criticism of the legislation of 1947 should be directed to the dualism that it introduced into the exercise of the Commonwealth general industrial authority. The demarcation of the respective provinces of Court and commissioners provided for is irrational. Why, for example, should the subject of annual leave be permitted to remain with the Court, while the similar matter of sick leave (a matter common to industry) is withdrawn from it? If standard hours are to be fixed by the Court, why should not the manner of their distribution over the week be a consideration for it as well? The quantum of the basic wage (assessed by the Court as the highest that industry can support) has a direct relation to the amounts paid out in the form of all other wages, because it is conditioned by those amounts (e.g., extra for afternoon, night and special shifts, margins for skill or responsibility, rates for juniors); how can the treatments be adequate, the fixations sound, and the relative wage structure throughout industry be preserved, unless the whole range of wages is within the one control? Far too much responsibility is cast upon those from whom no qualifications of any kind are legally demanded. The integral objection, however, goes to the *fact* of the unqualified dualism — the absence of provision for organic connection at any stage in the division of jurisdiction between the parallel authorities, both as regards the Court and the commissioners vis-a-vis, and with respect to the commissioners inter se. The legislators paid insufficient heed to the warning of Australian experience — the necessity for consistency in labor decisions and the need for ensuring that this is, to the greatest practical extent, rendered possible. The disregard is the more noteworthy because of the dissatisfactions and unsettlement caused by the Court's lapses in the war days, and by the conflicts between its decisions and those of the special authorities. No mention was made by the legislature, as a justification for emasculating the Court's arbitral jurisdiction, of the failures of the Court to observe a criterion that had been its mainstay in other years.

The labor situation in Australia at present is most unsatisfactory. The workers are discontented, the unions aggressive, employers resentful, awards and orders of conciliation commissioners and the special authorities often defied, while stoppages in the factories are frequent. There has been a decline in the respect that previously was extended to the industrial law, and in mutual good feeling in employer-employee relationships. Inter-union strife has intensified and the severity of union demarcation disputes has increased. A result

is an air of confusion and an absence of stability in industry.   In apportioning the blame for this condition political and ideological factors should not be overlooked, but to the condition the cleavage in the industrial authority and the imperfections in the industrial code are contributing in a substantial measure.   The conclusion is inescapable that no future is possible for a scheme of things that does not provide for adequate co-ordination and coherence of operation, and the ensuring of a body of decisions capable of satisfying the tests of uniformity and comparability.   A return should be made to the principle of one corporate authority fixed with a responsibility for the maintenance of industrial peace.   Little criticism is apparent in an arrangement by which, within that responsibility, each judge is made responsible for the conduct of specified industries; Australian experience has shown that, by that device, an arbitrator becomes familiar or intimate with the personnel, problems and technique of particular industries.   Working under the direction of the judges, and always in a supporting and subsidiary capacity, there could be conciliation commissioners, of a required number, much the same as previously was the case.   There should be no compromise with the principle of unity of ordered administration.

The ship of federal industrial regulation has slipped her moorings and is adrift.   There must be found for her a safe anchorage before irretrievable self-injury is sustained, and more serious damage inflicted upon industry.   The anchorage need not — in fact, clearly should not — be fully identical with that occupied before the démarche of 1947.   No one will quarrel with most of the objects of the legislation of that year.   Admittedly it was desirable that cumbersome technical preliminaries should be eliminated as much as possible in the interests of a more direct and speedy intervention in industrial disputes.   Admittedly the machinery of industrial regulation was overburdened with legalism, formal procedures and other artificial hindrances to expedition.   Undeniably there was room for a greater emphasizing of conciliation techniques in the scheme of industrial control.   But all reasonable reforms and desiderata in these directions are attainable within the framework of a unified system, without the failures that are inevitable when considerations of consistency are departed from.

Ultimately, however, it must be recognized that no system of industrial regulation introduced by the Parliament of the Commonwealth can evade, or make light of, the limitations imposed upon the legislative authority of that Parliament by the constitution.   Nearly all the charges of legalism that have been levelled against the Court

are based upon a failure to appreciate this cardinal fact: that a federal system of government is by nature and inevitably a legalistic system. Its instrumentalities can never escape the bounds imposed upon them by the constitutional limitations. It is both unreasonable and unjust to blame them for recognizing their unavoidable short-comings and their inherent weaknesses.

ORWELL DE R. FOENANDER.

THE UNIVERSITY OF MELBOURNE, AUSTRALIA

# INVESTMENT REPERCUSSIONS: A COMMENT

In a recent article in this *Journal*,[1] Dr. Lachmann discusses the phenomenon of investment in new capital resources making existing capital more productive. This implies that an entrepreneur making such an investment would maximize his profits by acquiring existing assets at their previous value, and himself enjoy the increase in their value. And he suggests that in Keynesian theory there is no place for capital complementarity. Yet there is surely no need for a special place for complementarity of this kind. As long as the entrepreneur has an opportunity of acquiring the existing capital goods at their previous price, he will clearly do so in maximizing profits, yet the increase in their productivity can be directly imputed to the new investment — that is, the increase in their productivity would indeed be regarded as part of the marginal efficiency of the new investment. The entrepreneur, then, would stop investing in the new resources only at the point where the increase in yield from the old investment, plus the yield of the new investment, was equal to the rate at which he could borrow. Yet there would still be the opportunity, in many cases, of expanding the quantity of capital resources of the old type, since the marginal productivity of these resources will have increased, even though the increase of capital value of the existing resources of this type were imputed to the new investment. If, then, the entrepreneur does indeed have the opportunity of acquiring existing assets, which are complementary to his new investment, at their old price, there will be no problem of a special nature — maximization of profits will ensure that investment continues beyond the range of significant complementarity. The entrepreneur will stop investing when the marginal efficiency of investment is once again falling.

The only reason, then, why complementarity should involve a special problem, would be when the entrepreneur *cannot* acquire, at their old price, existing assets. When he cannot prevent other people from enjoying increases in income imputable to *his* investment, he will stop investing before the marginal efficiency to the community of this investment is equated to the marginal cost. It is this type of investment which has no place in the Keynesian system — nor, surely, in any other general system. The division between the types of complementarity, being in terms of the entrepreneur's opportuni-

1. L. M. Lachmann, "Investment Repercussions," this *Journal*, November 1948.

ties of acquiring existing capital assets at their previous prices, the question arises as to why sometimes he cannot do so. In some cases he may be unable to acquire them because the gain is spread thinly over a large number of resources, and the supply curve of credit facing him may not be elastic. But the more important reason would surely be that the owners of the existing capital goods may anticipate and expect his investment, and would, therefore, only sell these goods at a price higher than the value in their previous uses. As soon as the owners of these existing resources *expect* complementary investment, they will be able to enjoy some of the external economies. So a Theory of Investment Repercussions, which Dr. Lachmann suggests is necessary, would involve assumptions as to the acuteness of holders of capital resources in expecting investment in other kinds of resources which happen to be complementary to their own. Of course, if there is an obvious case of this kind, then they will themselves invest in this new line of investment, in order to maximize their profits, unless they also are facing inelastic supply curves of credit. But in any case, it is surely improbable that this complementary investment could be completely concealed from existing capital owners, while their resources were bought at existing use value. But to hope for any general conclusions from a theory which uses as part of its data the foresight and knowledge of capital owners, when there is no a priori reason for expecting any uniformity between owners (except that those who are very short-sighted tend to go out of business, unless they are already monopolists) is surely a trifle optimistic.

Dr. Lachmann points out, of course, that the marginal efficiency of capital is explicitly defined with respect to " 'not the market-price at which an asset of the type in question can actually be purchased in the market, but the price which would just induce a manufacturer to produce an additional unit of such assets.' Prices of existing assets do not come within its scope. But if they do not come within its scope they cannot influence investment decisions, while in reality, undoubtedly, income and capital gains to be derived from the purchase of existing capital assets, complementary to the new investment are a very strong incentive to invest." The market-price at which an asset can be acquired, however, refers to assets in which the new investment is to take place. There is nothing incompatible with this definition of the marginal efficiency of capital, in the fact that income and capital gains can be made as a result of complementarity and that this adds to the incentive to invest, when we acquire existing assets *of a different kind* at their old market

prices. These market prices will necessarily affect the marginal efficiency of capital in the new line of investment.

That these investment repercussions take place is, of course, undoubted. Whether a generalization can be made about them, however, is another matter. Productivity of existing assets has changed, just as productivity changes when knowledge increases, or inventions are made. Yet one would not seek to construct a theory of invention, even though some inventions are "induced." I wonder, therefore, whether there is any more meaning in a concept of a Theory of Investment Repercussions.

DOROTHY HAHN.

UNIVERSITY OF BIRMINGHAM, ENGLAND

## REPLY

In her comment on my article[1] Mrs. Hahn makes two major points. In the first place, she criticizes my contention that, where old and new assets are complementary, the marginal efficiency of capital is not a meaningful concept. Secondly, if I understand her correctly, she denies the *logical possibility* of a Theory of Investment Repercussions.

Let there be a capital asset $A$ with an annual yield $a$. Suppose now that an investor contemplates the creation of a new asset $B$, the annual yield of which he estimates as $b$, but that he regards $b$ as an unsatisfactory yield. If, however, $A$ and $B$ are complementary assets, the annual yield on $A$ as a result of the creation of $B$ may rise from $a$ to $a + a'$. If our investor can buy $A$ at a price capitalizing $a$, he may find that a yield $a + a' + b$ on a capital outlay $A + B$ may be quite satisfactory after all. It is clear that it is the *combined* yield $a + a' + b$ on $A + B$ which will determine investment decisions. It is equally clear that it does not indicate the amount of new investment unless the ratio of old to new assets were always the same which it is not.

Mrs. Hahn feels that $a'$ should be imputed to $B$ and $a' + b$ regarded as the marginal efficiency of $B$. The argument is nugatory. What matters here is not what an observing economist would "impute," but what in reality will affect people's actions. She subsequently admits that how much of the yield increase $a'$ will actually accrue to our investor depends on the price at which the present owner of $A$ is willing to sell it. And his supply price depends on his own expectations, and not on any "imputation" on our part. If,

1. In this *Journal*, November 1948, pp. 698–713.

as Mrs. Hahn suggests, the *whole* yield increase $a'$, irrespective of to whom it actually accrues, is to be included in the marginal efficiency of $B$, the latter will not determine new investment. While, if only that part of $a'$ which is expected to accrue to the investor in $B$ is included, the marginal efficiency of $B$ becomes dependent on expectations of others than the investor. This conclusion leads us straight into the problem of inconsistent expectations which is the real issue at stake.

As regards $A$, Mrs. Hahn also makes the statement that " . . . the marginal productivity of these resources will have increased, even though the increase of capital value of the existing resources of this type were imputed to the new investment."[2] This is an astonishing contradiction in terms as, if the whole of $a'$ is to be imputed to $B$, the marginal productivity of $A$ cannot rise.

On the general question of the possibility of a Theory of Investment Repercussions Mrs. Hahn expresses grave doubts. "To hope for any general conclusions from a theory which uses as part of its data the foresight and knowledge of capital owners, when there is no a priori reason for expecting any uniformity between owners . . . is surely a trifle optimistic."[3] Such a view of the subject-matter of dynamic economics, expressed in 1949, is surely a trifle out of date. Mrs. Hahn thinks we can reach no "general conclusions" about a situation in which expectations are not uniform. But inconsistent expectations are the very bread and butter of dynamics. It was precisely in order to be able to deal with inconsistent expectations and their consequences that modern economists discarded equilibrium analysis in favour of process analysis. If Mrs. Hahn scornfully dismisses the prospect of a Theory of Investment Repercussions she must therefore mean the equilibrium type of theory in which all the conclusions are already implicit in the "data." And this raises a fundamental methodological issue.

That investment repercussions are a feature of economic reality, Mrs. Hahn does not deny. But she does not want us to study them, their uniformities of sequence, their causes and consequences. Why? Because if we use as part of our data "foresight and knowledge" which are not "uniform," the results of action will not already be implicit in our data. But the Theory of Investment Repercussions is a dynamic theory. It deals with real processes, not with fictitious states of rest. As a process unfolds, people gradually acquire new knowledge about each other as well as about the objective facts of

2. p. 430.
3. p. 431.

the situation confronting them, and, our original "data" will not long remain what they were. Of Mrs. Hahn's objection we may say therefore that "the tautological method which is appropriate and indispensable for the analysis of individual action seems in this instance to have been illegitimately extended to problems in which we have to deal with a social process in which the decisions of many individuals influence one another and necessarily succeed each other in time."[4]  Professor Hayek has shown that the transmission of knowledge, initially by no means "uniform," constitutes the real empirical content of economic and social processes.  That one of his pupils of all people, ignoring the fundamental distinction between the tautological method of equilibrium analysis and the realistic method of process analysis, should want to rule out a promising field of study for the sole reason that it is unlikely to fit into the Procrustean bed of the Pure Logic of Choice adds a fine touch of irony to the pleasures of Mrs. Hahn's little excursion into the methodology of the social sciences.

But strangest of all is that Mrs. Hahn, in her eagerness to nip the Theory of Investment Repercussions in the bud, appears to have overlooked the fact that, albeit in a crude and naive form, such a theory is already in existence: the "stagnation thesis" which asserts that the effect of present investment on future investment decisions is always bound to be detrimental.  With this theory we may agree or disagree, or we may suspend judgment until more evidence is available. There is only one thing we may not do: deny its existence, bar its critical discussion and elucidation, and "wonder whether there is any meaning in a concept of a Theory of Investment Repercussions."

<div align="right">L. M. LACHMANN.</div>

UNIVERSITY OF THE WITWATERSRAND, SOUTH AFRICA

4. F. A. Hayek, *Individualism and Economic Order*, p. 93.

# RECENT PUBLICATIONS

ANSHEN, MELVIN. An Introduction to Business. New York: Macmillan. 1949. pp. 672. $5.00.

BETTELHEIM, CHARLES. Esquisse d'un Tableau Économique de l'Europe. Paris, France: Domat Éditions. 1949. pp. 356.

BEYEN, J. W. Money in a Maelstrom. New York: The Macmillan Company. 1949. pp. 221. $3.25.

BOGUE, DON J. The Structure of the Metropolitan Community: A Study of Dominance and Subdominance. Ann Arbor, Michigan: University of Michigan. 1949. pp. 210. $2.00.

BOWEN, HOWARD R. Toward Social Economy. New York: Rinehart & Company, Inc. 1949. pp. 336. $3.00.

BOWER, PENELOPE A. Balance of Payments of Nigeria in 1936. Oxford, England: Basil Blackwell. 1949. pp. 86. 10s. 6d.

BOWMAN, MARY JEAN AND BACH, GEORGE LELAND. Economic Analysis and Public Policy. New York: Prentice-Hall, Inc. 1949. pp. 931. $5.00. [Second edition.]

BOYD, RALPH L. AND DICKEY, ROBERT. Basic Accounting. New York: Rinehart & Company, Inc. 1949. pp. 687. $5.00.

BRIGHT, ARTHUR A., JR. The Electric-Lamp Industry. New York: Macmillan. 1949. pp. 526. $7.50.

BUCK, A. E. Financing Canadian Government. Chicago, Illinois: Public Administration Service. 1949. pp. 367. $5.00.

BURNS, ARTHUR F. Wesley Mitchell and the National Bureau. New York: National Bureau of Economic Research. 1949. pp. 107.

BURNS, EVELINE M. The American Social Security System. Boston, Mass.: Houghton Mifflin Company. 1949. pp. 460. $4.50.

CAHN, FRANCES T. Federal Employees in War and Peace. Washington 6, D. C.: The Brookings Institution. 1949. pp. 253. $3.50.

CHANG, PEI-KANG. Agriculture and Industrialization. Cambridge, Mass.: Harvard University Press. 1949. pp. 270. $5.00.

CONYNGTON, THOMAS AND BERGH, LOUIS O. Business Law. New York: The Ronald Press Company. 1949. pp. 866. $5.00.

DAVIS, PEARCE. The Development of the American Glass Industry. Cambridge, Mass.: Harvard University Press. 1949. pp. 316. $6.00. [Harvard Economic Studies.]

DICKEN, SAMUEL NEWTON. A Regional Economic Geography. Boston: D. C. Heath and Company. 1949. pp. 516. $5.50.

DIMOCK, MARSHALL EDWARD. Business and Government. New York: Henry Holt and Company. 1949. pp. 840. $4.75.

DYKSTRA, GERALD O. AND DYKSTRA, LILLIAN G. Business Law: Text and Cases. New York: Pitman Publishing Corporation. 1949. pp. 1181.

GENERAL EDUCATION BOARD. Annual Report, 1947-48. New York: Rockefeller Foundation. 1949. pp. 304.

GLOVER, JOHN G. Business Operational Research and Reports. New York: American Book Company. 1949. pp. 299. $4.00.

GOLDENWEISER, E. A. Monetary Management. New York: McGraw-Hill Book Company. 1949. pp. 175. $2.75.

GRIFFIN, CLARE E. Enterprise in a Free Society. Chicago, Illinois: Richard D. Irwin, Inc. 1949. pp. 573. $5.00.

GULICK, LUTHER. A New American Policy in China. Washington, D. C.: National Planning Association. 1949. pp. 35. 50 cents.

435

HARPER, F. A.   Liberty: A Path to its Recovery.   Irvington-on-the-Hudson, New York: The Foundation for Economic Education.  1949.  pp. 160.  $1.00.

HARRELL, THOMAS WILLARD.   Industrial Psychology.   New York: Rinehart & Company, Inc.  1949.  pp. 462.  $4.25.

HARRIS, SEYMOUR E.   Economic Planning.   New York: Alfred A. Knopf, Inc. 1949.  pp. 577.  Trade $6.00, Net $4.50.

HAWLEY, AMOS H.   The Population of Michigan 1840 to 1960: An Analysis of Growth, Distribution, and Composition.  Ann Arbor, Michigan: University of Michigan Press.  1949.  pp. 116.  10 cents.  [Michigan Governmental Studies, No. 19.]

THE HOOVER COMMISSION.   The Hoover Commission Report on Organization of the Executive Branch of the Government.  New York: McGraw-Hill Book Company, Inc.  1949.  pp. 524.  $3.75.

HYDERABAD STATE GOVERNMENT.   The Hyderabad Government Bulletin on Economic Affairs.   Hyderabad, India: The Economic Adviser, Finance Department.  1949.  pp. 145.

JONES, FRED M.   Principles of Retailing.   New York: Pitman Publishing Corporation.  1949.  pp. 651.

LEVY, JEROME AND LEVY, S. JAY.   The Capitalist System.   New York: New Economic Library.  1949.  pp. 20.  25 cents.

MACDONALD, PHILIP.   Practical Exporting.   New York: The Ronald Press Company.  1949.  pp. 355.  $4.00.

MANSFIELD, HARVEY C.   A Short History of O.P.A.   Washington, D. C.: Office of Temporary Controls, O.P.A.  1949.  pp. 332.  55 cents.

MILLS, FREDERICK C. AND LONG, CLARENCE D.   The Statistical Agencies of the Federal Government.  New York: National Bureau of Economic Research, Inc.  1949.  pp. 201.  $2.00.

MINISTRY OF LABOUR GOVERNMENT OF INDIA.   The Indian Labour Year Book, 1947–48.  Simla, India: The Ministry.  1949.  pp. 296.  9 shillings.

MOULTON, HAROLD G.   Controlling Factors in Economic Development.   Washington, D. C.: The Brookings Institution.  1949.  pp. 397.  $4.00.

NATIONAL INDUSTRIAL CONFERENCE BOARD.   Social Security Almanac.   New York: The Conference Board.  1949.  pp. 112.

OWEN, WILFRED.   Automotive Transportation Trends and Problems.   Washington, D. C.: The Brookings Institution.  1949.  pp. 154.  $2.00.

PEACH, W. NELSON AND KRAUSE, WALTER.   Basic Data of the American Economy.  Chicago, Illinois: Richard D. Irwin, Inc.  1949.  pp. 228.  $2.00.  [Revised Edition.]

PETTENGILL, ROBERT B.   Price Economics.   New York: The Ronald Press Company.  1949.  pp. 483.

PIETRI-TONELLI, ALFONSO DE.   Un Sistema di Logodiagrammi Tipici.   Milan, Italy: Cedan.  1949.  pp. 67.

PIGOU, A. C.   Employment and Equilibrium.   New York: The Macmillan Company.  1949.  pp. 283.  $4.00.

PRATHER, CHARLES L.   Money and Banking.   Chicago, Illinois: Richard D. Irwin, Inc.  1949.  pp. 692.  $5.00.

RESEARCH AND PLANNING DIVISION, E.C.E.   UNITED NATIONS, DEPARTMENT OF ECONOMIC AFFAIRS.  Economic Survey of Europe in 1948.  Geneva, Switzerland: The United Nations.  1949.  pp. 288.  $2.50.

RUGINA, ANGEL.   Geldtypen und Geldordnungen.   Stuttgart, Germany: W. Kohlhammer.  1949.  pp. 352.  DM 20.

SHACKLE, G. L. S.   Expectation in Economics.   Cambridge, England: Cambridge University Press.  1949.  pp. 146.  $3.00.

SOMERS, HAROLD M.   Public Finance and National Income.   Philadelphia: The Blakiston Company.  1949.  pp. 540.

TAX INSTITUTE SYMPOSIUM.   Income Tax Administration.   New York: Tax Institute.  1949.  pp. 449.  $6.00.

UNITED NATIONS. Technical Assistance for Economic Development: An Expanded Plan. New York: Columbia University Press. 1949. pp. 328. $2.50.

UNIVERSITY OF ILLINOIS. Current Economic Comment. Urbana, Illinois: College of Commerce and Business Administration. Feb., 1949. pp. 63. [Formerly Opinion and Comment.]

U. S. DEPARTMENT OF COMMERCE, BUREAU OF THE CENSUS. Historical Statistics of the United States, 1789–1945. Washington 25, D. C.: The Bureau. 1949. pp. 363. [A Supplement to the Statistical Abstract of the United States.]

WAITE, WARREN C. AND CASSADY, RALPH, JR. The Consumer and the Economic Order. New York: McGraw-Hill Book Company. 1949. pp. 440. $4.50.

WARE, NORMAN. Wealth and Welfare. New York: William Sloane Associates, Inc. 1949. pp. 231. $1.90.

WEBER, ADOLF. Markt Wirtschaft und Sowjet-Wirtschaft. München: Richard Pflaum. 1949. pp. 465.

# QUARTERLY JOURNAL ANNOUNCEMENTS

**Student Subscriptions.** The annual rate to students is increased beginning July 1 to $3.00 per year. This change is made necessary by a recalculation of marginal costs.

**Still Available: The A & P Case.** There are still a few copies on hand of our May issue containing M. A. Adelman's "The A & P Case: A Study in Applied Economic Theory" for which there has been considerable demand. The ideas of the Antitrust Division as developed in the case against the New York Great Atlantic & Pacific Tea Company during 1945 and 1946 have prevailed, and will influence public policy and business conduct. Thus, they deserve examination. In his article, Mr. Adelman explores the economic theory of the prosecution and its implied economic policy in relation to the business methods of the A & P Company in selling goods, integration of manufacturing and retailing, and buying. If you have not yet seen this article, order your copy of the *Quarterly Journal of Economics* now. It appears in the May, 1949, issue. Order from the

# THE

# QUARTERLY JOURNAL

# OF ECONOMICS

## ISOLATIONISM IN ECONOMIC METHOD[1]

### SUMMARY

I. The isolationism of the orthodox, Marxist, and historical schools in their concepts of economic science, 439; in their distribution theory, 441. — II. Problems of definition and classification, 442; distinction between "ultimate" and working problem, 444. — III. Formulation of the ultimate problem of distribution theory, 445. — IV. The working problem of the orthodox or maximization school, 447; of the Marxist school, 451; of the historical school, 457. — V. Components of isolationism, 465. — VI. Origins of isolationism, 468. — VII. Conclusion, 471.

I

A large proportion of economists throughout the world today entertain a "solipsistic" attitude toward the field of economic inquiry: they act as though only they, and other economists sharing their point of view, occupy the field known as "economic science." To western members of the profession it should come as no surprise to be told that economists on the other side of the "curtain" engage in this practice — that they identify economic science as such with the doctrines on this subject expounded a century ago by Karl Marx and Friedrich Engels, as developed and interpreted by Vladimir I. Ulianov and Joseph Djugashvili, and dismiss all other theories as "bourgeois apologetics," "vulgar political economy," or "capitalist fabrications." A readily available example of the positive side of this procedure is the widely discussed article of a few years ago by the editors of a Soviet journal dealing with the "Teaching of Economics in the Soviet Union."[2] Throughout the article its authors, who, so far as we know, have received no official reprimand or other punishment for doing so, consistently identify the terms "economics" or "political economy" with "Marxian economics" or "Marxist political economy." Their suggestions for improving the teaching of economics in Soviet schools

1. The author is indebted to Professors J. M. Clark, F. K. Mann, and J. A. Schumpeter, and to Dr. O. H. Taylor for reading the manuscript and offering helpful criticisms and suggestions.
2. *Pod Znamen Marxizma*, Nos. 7–8, 1943, R. Dunayevska, transl., *American Economic Review*, XXXIV (1944), 501–30.

are intended ostensibly to cleanse economics of certain bourgeois errors that have unfortunately captivated some of their teaching comrades, and to restore economics to its pure, i.e., Marxist, form. Whether or not their suggested changes would reorient economics courses in the Soviet Union toward pristine Marxism is irrelevant here; more significant is the fact that any Soviet economists who attempted to answer them would be obliged to use the same formula, a formula which rests on the assumption that "economics" equals "true economics" and that "true economics" equals "economics of the Marxist school."

This "oyster" view of the field of economic inquiry, in the form of a school centrism, unfortunately is not confined to the Soviet Empire and its representatives abroad. In the western world today there is another school of economic theory which has been so sanctified by traditional acceptance and has remained so impervious to attack by its rivals here (including the Marxists), that it is identified, in the vocabulary of most of our economists, with economic theory as such. The members of this school regard themselves not as a school but as the only practitioners of the economic profession; perhaps because they are more genial, polite, subtle, or smug than the Soviet economists, they dismiss critics less for being "liars" or "fakers" than for being non economists — i.e., "sociologists," "psychologists," or "historians." To take but one example, an eminent contemporary leader of this school defends what he calls "Economics," "Economic Science," "Economic laws," "the propositions of Economics," etc. against those whom he labels "the detractors" and "the critics of Economics"; he solves problems and enunciates verities not from the viewpoint of himself or his school but "from the point of view of Economic Science."[3] This attitude, despite important changes in economic doctrine in recent years, particularly those dealing with what we may call "limited competition," is very widespread; perhaps the majority of western economic theorists, in formal or informal discussion with one another today, use the term "economic theory," "pure theory," or simply "theory" to refer to a technique of economic analysis which is considerably short of encompassing the range of investigation literally implied by the term. Like the Marxists (though less narrow), these economists assume an identity between "economics" and "true economics" but identify the latter with what has been called by its critics

3. Lionel Robbins, *The Nature and Significance of Economic Science* (2d ed.; London: Macmillan & Co., 1935), pp. 25, 27, 45, 72, 83, 85, 104, 106, 123. Such passages by orthodox economists or Marxists often sound like variations on the ancient theme: "It is not I, a human sinner, that addresses you, but God, speaking through the mouth of His prophet."

and its less solipsistic members "orthodox economics" or "the ortho-
dox school of economics."

Just as the orthodox economists and the Marxists, on the whole,
have recognized each other's myopia more than their own, so the
members of a third school, the historical school, who for many years
have objected to the narrowness of the orthodox, and to a lesser extent
of the Marxist, interpretations of the field of economic inquiry,[4] have
developed their own variety of the same disease.  One of the founders
of the United States branch of this school, Thorstein Veblen, had too
Darwinian a sense of historical continuity to exclude from a genus any
of its evolving series of species, no matter how antediluvian or anach-
ronistic.  He therefore classified the orthodox, Marxist, and German
historical schools within "economics" and to some extent (in their
latest forms) within "economic science."  But he insisted that these
schools belonged, on the whole and in varying degrees, either to the
species of "animistic metaphysics" or to that of "normalistic taxonomy"
and as such were forbears rather than full-fledged members of the
species which he described as "modern science" (i.e. "evolutionary,"
"Darwinian," "matter-of-fact") of which he was a spokesman.[5]

The following analysis will attempt to show, first, that this
*l'univers c'est moi* attitude which most members of the three major
traditional schools share in common (each from his or its own vantage
point) toward the scope of "economics," "economic science," or
"modern economic science" is a manifestation of a similarly exclusive
or isolationist attitude toward the concrete problems of economic
inquiry, exemplified here by the problem of distribution theory.  A
second objective will be to analyze the isolationism into its component
parts and to suggest briefly some of the similarities and differences
among the schools, both in their degrees of narrowness and in its
origins.  Finally, an effort will be made to explore the extent to which
the isolationism is a methodological advantage, a necessary evil, or a
remediable vice.

4. E.g., Walton Hamilton in "The Place of Value Theory in Economics,"
*Journal of Political Economy*, XXVI (1918), 217–45 and 375–407 protested the
orthodox identification of "economic theory" with "value theory."
5. "Why is Economics not an Evolutionary Science?", "The Preconceptions
of Economic Science: I–III," "Gustav Schmoller's Economics," and "The Socialist
Economics of Karl Marx: I–II," this *Journal*, XII–XIV (1898–1900), XVI
(1901), XX (1906), and XXI (1907), reprinted in his *The Place of Science in
Modern Civilization* (New York: The Viking Press, 1919), esp. pp. 57–8, 67–8, 70,
81–5, 109–10, 144–7, 171–9, 269–78, 413–18, and 432–56.  Veblen's classification
of scientific schools and stages is not very precise, but that, as he would say, is a
mere matter of taxonomy.

## II

Several preliminary matters must be examined first.

(1) This article examines the isolationism of the traditional schools as it is exemplified in their treatment of the problem of distribution theory. The term "distribution theory," however, is interpreted here in a broad sense which covers the study of the comparative net income shares of different transactors in a given time period or a succession of time periods of the capitalist economy. Since net income shares are a function of relative prices and relative sales volumes (or production volumes) of commodities bought and sold, distribution theory, so interpreted, includes the portion of price theory (traditionally referred to as "value theory") and of production theory which deals with comparisons among different coexisting transactors or groups of transactors at a given moment or during dynamic (including cyclical and secular) movements. The term distribution theory, therefore, includes a large portion of what has been traditionally regarded as economic theory; it excludes the static and dynamic theories of aggregates and averages for the economy as a whole (such as national income, product, savings, and investment; and cyclical and secular changes in these or in price levels).

(2) The problem of distribution theory here is a descriptive problem. Ethical and pragmatic issues, though important as incentives and consequences of distribution theories, are relevant only indirectly.

(3) In a comparative study of economic theories, some method of classifying the theories and the economists is necessary. Well-known difficulties include (a) the overlapping, change, and interaction among many of the phenomena described by the theories; (b) the differences in inclusiveness and emphasis among the theories, which causes them to be classed under more than one category; (c) the multiplicity of ideas generally contained in the system of theory of a single economist; and (d) an understandable dislike on the part of many economists of being classified at all, or at least of being put into categories in which the company is uncongenial, their special contributions are overlooked, and what they regard as the basic divisions in doctrine are obliterated. Let it be said at once that the author of the present article recognizes the partly but unavoidable arbitrary nature of his categories and of his parcelling out of theories among them. He apologizes for having to put into the same compartments the theories of economists, some or all of whom may find the company uncongenial. He is interested in economists here not as personalities but as convenient symbols of their doctrines. And finally, he fully

expects that he and his own theories will be categorized by his erst-while subjects, probably into pigeon-holes which he too dislikes. The significant issues then become: (a) are the category distinctions in question valid and important, and (b) do the theories approximately fit the categories in which they have been placed?

(4) The major type of category used here in classifying theories, or their authors, is that of "schools." A school may be defined as a class of theorists who either (a) recognize a basic identity in their ideas and perhaps have entered into a formal or informal organization to develop and propagate those ideas, or (b) possess ideas which to an external observer are in some important respect identical, whether or not the individual members regard themselves as members or the identity as important. In the present essay the term school is used in the latter, objective sense. Each school, moreover, is to be distinguished from others by the method of "ideal types" or "models" wherever possible; i.e., emphasis is to be on the purest form of each, without forgetting that the schools are thereby simplified for methodological purposes.

(5) The major traditional schools of economic theory dealt with here are the orthodox, Marxist, and historical schools, with certain terminological and other qualifications to be presented below. No implication is intended that this classification exhausts the field or is the only one possible. But these schools, defined broadly, are the ones which most economists would probably recognize as the most important and influential schools of the past hundred years, at least so far as the problem of distribution theory is concerned.

(6) The term "problem" is itself a problem requiring careful scrutiny. A clear statement of a problem, it is often asserted, is half its solution. But scientific writers do not always clarify to their readers or to themselves the exact object of their search. Examination of the writings of economists reveals a large number in which the problem is more implicit than expressed; or, if elaborated, is more ambiguous than clear; or, if clearly stated at the beginning of a study, is a different problem from the one finally answered. Thus it is often difficult to decide, after perusing an economic treatise, just what was the problem actually motivating the author. The confusion sometimes arises not from the difficulties in the economic subject matter and in the arts of writing and reading but from imperfections in methodological self-consciousness on the part of the inquirer.

The problems under consideration here are theoretical rather than empirical, in form of presentation if not in historical order of discovery. The conditions influencing the outcome of a scientist's

inquiry are (a) the scope of his subject matter, (b) his objective, (c) his data (i.e., general and particular statements of fact concerning the subject matter which he accepts as accurate), and (d) the methods he intends to follow in proceeding from (c) to (b).[6] In an empirical inquiry, with any success, the scientist's data at the end of his inquiry will differ at least in quantity from the data with which he began; but in a theoretical investigation the data remain unchanged, unless the scientist is illogical or develops his conclusions by the method of successive approximation (as Marx does, for example).

Should all the conditions for the solution of a theoretical problem be stated explicitly? The answer, of course, is yes. If the term problem is defined to include all these conditions, the solution, unless illogical, becomes a tautology, being read off in the same manner as the solution performed by a calculating machine; a clear statement of the problem becomes its entire solution. Far from being undesirable, this tautology[7] is an ideal almost impossible of attainment. Usually the scientist regards his subject matter and objective as parts of his problem but does not state clearly his assumptions concerning data and method; in this event, two scientists may regard themselves as investigating an identical problem when the conditions respectively postulated for their investigations are so different as to impel them to different solutions (unless their chains of reasoning logically or illogically happen to intersect at the same terminus). One may conclude that they are in effect examining different problems.

In order to avoid these and other difficulties, a distinction is hereby suggested between an inquirer's "ultimate problem" and his "working" or "operational problem." An ultimate problem is here taken to mean the problem (or complex of problems) coextensive in breadth with the entire range of the subject matter from the viewpoint or objective of the science or branch of the science in question; a working problem is that same problem (or complex of problems) narrowed down by the major assumptions of data and method presupposed or examined by the individual inquirer or school. The ultimate problem is broad enough to enable all inquirers or schools within the science to

6. The distinction between these four conditions should not be drawn too rigidly.

7. It is a tautology only in relation to the fully stated premises, like any other system of syllogisms. The empirical collection of data for use as premises then appears to constitute the progress of science. But this overlooks the human effort and skill involved in formulating and ordering premises and in drawing correct inferences from those premises, even though, after the problem is solved, the solution is seen to have been "implicit." It also overlooks the fact that the significant premises in which the solution is implicit are usually discovered after the solution has been arrived at.

stand on the same footing while the logical answers to a working problem are essentially alike. An important part of a scientist's inquiry is to examine to what extent his solution of the narrower problem is a solution of the broader one, which requires that he examine carefully the studies made by those with alternative working assumptions. His own working problem broadens beyond his own working assumptions to the extent that he seriously examines alternative working assumptions and dismisses them on more substantial grounds than mere incompatibility with his own.

To determine a scientist's ultimate problem external observers must examine as a whole the field of inquiry in which he appears to be operating. To determine his working problem, they must examine his positive operations and the restrictions on his operations which he appears to be following within that field. In determining both, observers must, of course, use his statements of his problem as part of the evidence.

## III

The ultimate problem of distribution with which the schools are concerned is easier to ascertain than is their working problem. We may begin with J. M. Clark's useful formulation: "The central problem of distribution in economic theory may be defined as the analysis of the forces which under free exchange govern the division of the product of industry between those who perform different functions or supply different factors. The shares may, however, be differently distinguished: according to the contractual arrangements under which the proceeds are received, according to the persons receiving them and according to the underlying functions or factors which constitute the sources of the incomes. Thus there are contractual distribution, personal distribution and functional distribution. Although these distinctions are generally made in economic textbooks and may be useful for didactic purposes, the main body of economic theory has for a long time concerned itself with functional distribution only."[8]

This definition was formulated for other purposes, but may be converted, with some changes and qualifications, into a definition of the ultimate problem in the sense used here. First, the term "free exchange" is subject to misinterpretation; whether free exchange (in at least two different interpretations of the term) actually exists in our economy or not, is a fundamental issue between certain schools of

8. "Distribution," *Encyclopaedia of the Social Sciences*, Vol. V (New York: The Macmillan Co., 1931), p. 167; reprinted in The American Economic Association, *Readings in the Theory of Income Distribution* (Philadelphia: The Blakiston Co., 1946), p. 58.

economic theory and therefore must not be prejudged in the phrasing of the problem. The monopolistic and imperfect competition economists deny in individually varying degrees that the present western economy is one of free exchange in the sense of "unlimited competition." The Marxists deny that workers in capitalism conduct free exchange, except in choosing between starvation and exploitation; on the other hand, they affirm that free exchange exists between state enterprises and individuals in the Soviet Union.[9] Yet all would probably agree that a common subject matter of their inquiry is an economic system in which (a) individuals or groups own diverse amounts of private property in natural resources, capital, and consumer goods; (b) individuals or groups dispose of their property or labor and acquire goods or services by making contracts with other private individuals; (c) in making contracts, individuals generally have more than one alternative, though not necessarily in the same market; (d) the terms of the contracts are agreed upon, whether reluctantly or not, by both sides; (e) the contracts are legally enforced; and (f) individuals who do not agree upon terms may refrain from making contracts. The opposite type of exchange system is one in which the provisions of transactions and the alternatives of transactors are fixed by some one authority — probably a government (whether democratic or dictatorial in form), possibly a single transactor with a comprehensive monopoly (who then becomes an economic dictator).

No completely satisfactory terms are available to describe and distinguish the two exchange systems. The one will be called a "centralized" or "authoritarian market system" (which is conceptually compatible with democratic control over the authority); the other, to avoid the emotional associations of the word free, will be called a "pluralistic" or "private market system" (which is conceptually compatible with monopolies in individual markets). Each of the two types of exchange systems shades into the other and neither is found in absolute form. An economy in practice may contain within itself numerous economic systems or sectors, taking its individual character from the sector which is dominant. The subject matter of distribution theory consists in the "private market system," wherever it may be found. This system constitutes a large portion but by no means the entirety of the modern capitalist economy.[1] It also has con-

---

9. E.g., Maurice Dobb, *Soviet Economic Development since 1917* (London: Routledge & Kegan Paul, Ltd., 1948), p. 350.

1. In an unpublished essay, the author estimated that in 1940, of the numerous sectors of the United States economy, including government services and employment, private charity, illegal modes of acquisition (such as racketeering), cooperatives, and production for own use, the private market sector, though still

stituted a smaller portion of most social economies known to us,[2] and may even be found in the Soviet Union today.[3]

Secondly, the ultimate problem of distribution theory cannot be restricted to functional distribution in the traditional sense because, as the quoted definition implies, this is a narrower objective than the one occupying the entire body of distribution theorists. That economists increasingly have been recognizing distributive divisions other than the traditional ones requires little substantiation. (See pp. 449–450 below.) The distinction between "functional" and "personal" distribution may be drawn in more than one way, and may or may not be a "didactic" distinction; this is one of the issues among the several schools and should not be prejudged in stating the problem common to all of them. (See pp. 464–465 below.) Professor Clark's definition of the "central problem" for "the main body of economic theory" thus needs further revision as the ultimate problem for the entire body of economic, or at least distributive, theory. The latter is concerned with distribution among economic groups or classes of all sorts, whether or not functional in nature and source, in so far as the distribution is accomplished via the market.

The ultimate problem of distribution theory, therefore, may be defined as the study of the influences which govern the division of the total product into net income shares among the various social groups and classes through the process of exchange in a private market system. This formulation of the problem, it is hoped, enables all schools of distribution theory to stand on the same footing. The one important objection which may be urged against it, however, is that it phrases the problem so broadly that it provides little or no guide as to how it may be solved. For such a guide we must turn to the "working" or "operational" problems of the orthodox, Marxist and historical schools.

## IV

The orthodox school may be delineated in several different ways, depending on how broad a coverage is granted to it. To abstract the common features of a school that has been constantly evolving since

the largest, accounted for less than half the total income of the economy's inhabitants. The computations, of course, were extremely rough.

2. See, e.g., the Old Babylonian (earliest known) legal codes, approximately 2000 B.C., of Eshnunna (*New York Times*, Late City Edition, March 7, 1949, pp. 1 and 13) and Hammurabi (Chilperic Edwards, *The Hammurabi Code: and the Sinaitic Legislation*, 3d ed.; London: Watts & Co., 1921, p. 25, §§ 108 and 111; pp. 75–6 note). Also Richard Thurnwald, *Economics in Primitive Communities* (London: Oxford University Press, 1932), pp. 141 and 159.

3. Maurice Dobb, *op. cit.*, p. 285.

its origin presents a never-ending and progressively more difficult task. For present purposes the orthodox school will be conceived broadly, to encompass among other sub-schools or branches the physiocratic, classical, marginal utility-productivity, neoclassical, general equilibrium, and the contemporary monopolistic and imperfect competition branches. Although present day theorists of this school are attempting increasingly to handle problems of dynamic theory in combination with the traditional static problems of value and distribution,[4] it will be convenient here to regard these studies, to the extent that they contradict the traditional assumption of self-adjustment, as excursions outside the orthodox fold, and, to the extent that they attempt to explain cyclical or secular aggregates, as outside distribution theory. Similarly, the efforts of some economists to bring the study of a socialist economy under the orthodox technique[5] will be excluded.

Most orthodox economists, at least from Carl Menger on, have regarded economics as the science which studies the maximization of human ends by the minimum expenditure of alternative scarce means. In application to a private market system, they regard economics as the study of the self-equilibrating influences which operate in private markets to accomplish and maintain a maximization of net income for each independent transactor consistent with that of the other transactors under the given environmental and other conditions.

The "maximization equilibrium theory" has had a long history. The doctrine of maximization alone in some form or other is, of course, an ancient one.[6] Numerous terms have been devised to describe this motive, or assumption, including not only the earlier phrases "self interest," "egoism," "hedonistic" or "felicific calculus," and "economic man," but also the more recent "principle of substitution," "principle of least means," "imputation technique," "economizing process," or the term "maximization" itself. The origin of the maxim-

4. E.g., J. M. Keynes, *The General Theory of Employment, Interest and Money* (New York: Harcourt, Brace and Company, 1936); J. R. Hicks, *Value and Capital* (Oxford: Clarendon Press, 1939), Part III; Moses Abramovitz, *Price Theory for a Changing Economy* (New York: Columbia University Press, 1939); and Paul A. Samuelson, *Foundations of Economic Analysis* (Cambridge: Harvard University Press, 1948), Part II.

5. Such as Oskar Lange and Fred M. Taylor, *On the Economic Theory of Socialism* (Minneapolis: The University of Minnesota Press, 1938) or A. P. Lerner, *The Economics of Control* (New York: The Macmillan Co., 1944).

6. E.g.: "the very thing which we all of us desire, — I mean in having a greater amount of pleasure and less of pain during the whole of life." (Plato, *Laws*, transl. by B. Jowett (New York: Random House, 1937), 5.733. See also *Republic*, 2.359 B.) Or, needless to add: "The love of money is the root of all evil." (*The Bible: Timothy* 1.6.10.)

ization *equilibrium* notion, though the latter is implicit in certain ancient and medieval European writings, probably should be attributed to the early critics of mercantilism, such as Dudley North.[7] The doctrine has, of course, been given much greater precision in recent years.[8]

An examination of the differences and refinements made in successive interpretations of the maximization equilibrium hypothesis during the development of the orthodox school and a criticism of the hypothesis itself lie outside the scope of this paper. Since the term "orthodox school" has been defined in various ways during the course of its evolution and since the term "orthodox" is too abstract, imprecise, and "value-loaded" for scientific use, it is here proposed to call the school under discussion the "maximization equilibrium school" or more briefly the "maximization school."

The maximization school's interpretation of the problem of distribution theory is simply one application of its problem of economic theory in general. Modern members of this school indicate in practice and occasionally assert explicitly that they have gone beyond the traditional divisions of economic theory (production and distribution; wages, interest, rent, and profits), replacing them by analyses of maximization equilibria for various types of assumed conditions. "We no longer enquire concerning the causes determining variations of production and distribution. We enquire rather concerning the conditions of equilibrium of various economic 'quantities,' given certain initial data, and we enquire concerning the effects of variations

7. *Discourses upon Trade* (1691).
8. See, for instance, George J. Stigler, *The Theory of Price* (New York: The Macmillan Co., 1946), pp. 12–13, 26; Joan Robinson, *The Economics of Imperfect Competition* (London: Macmillan and Co., Ltd., 1933), pp. 6, 15, 16, 212; Paul A. Samuelson, *Foundations of Economic Analysis*, p. 5; John von Neumann and Oskar Morgenstern, *Theory of Games and Economic Behavior* (Princeton: Princeton University Press, 1944), pp. 34, 44, 45. Stigler comments that "the principle of the maximum is not strictly necessary, for an economic problem would arise if individuals sought to minimize the attainment of their ends, or if they sought 65 per cent of the maximum possible attainment of the ends, or any other specific objective. But it is enormously convenient and, on the present plane of generality, very realistic to assume that a maximum fulfillment of ends is sought" (p. 13). But if individuals sought a minimum, 65 per cent of a maximum, or any other objective, these would become the ends to be maximized, e.g., on the part of abstemious Stoics, hair-shirted monks, or "give-away-checkers" players. Thus this form of assertion of the maximization hypothesis makes the principle of the maximum not only "enormously convenient" and "very realistic" but logically incontrovertible — one can deny it only by contradicting one's self. This is because the maximization principle so defined is a tautology. If maximization is defined in terms of some immediate, objectively verifiable end, such as the attainment of money, as is done by other representatives of the school, the principle may be deprived of its tautological character, but then becomes decidedly controvertible.

of these data."[9]  Such assertions should not be taken to mean that
the maximizationists do not attempt to explain, or do not think they
can explain, the distribution of income in general or among the tradi-
tional functional classes in particular.  Rather, they mean that their
essential concern is the broader one of finding maximization equilibrium
solutions for markets of given characteristics, from which solutions
anyone so inclined may draw inferences concerning relative pro-
duction and sales volumes, prices, or income-shares of transactors,
whether divided according to traditional categories or not.  Although
the maximizationists still discuss distribution among the traditional
functional classes,[1] they imply that these classes are used for illus-
trative purposes only, and that their technique of analysis can be
applied equally well to other class divisions.  Nevertheless, whatever
class divisions they do use are of a functional nature.

To summarize — the working problem of distribution theory for
the maximization school is to seek a determinate, self-preserving set
of net income maxima among the functional classes of transactors in a
private market system under various assumed market and environ-
mental conditions.  These maxima are self-preserving or stable in the
sense that they constitute the highest net income for each independent
transactor (person, enterprise, or association) consistent with that of
every other independent transactor under the assumed conditions.
Given a stable maximization equilibrium, any alteration in a trans-
actor's contract terms, which may seem to increase or decrease his
net income, would set in motion the self-correcting influences and the
previous equilibrium would be restored, if the transactors' preference
scales based on the assumed constants such as the natural environ-
ment, tastes, technology, laws, population, number of firms, etc.
remained unchanged or if a recontracting assumption were made.

This working problem is narrower than the ultimate problem
(p. 444) with respect to at least four sets of assumptions: (1) certain
types of phenomena remain fixed or at least are not changed by and
do not react upon the processes under analysis; (2) the transactors
maximize their net incomes under the given conditions; (3) market
forces achieve and preserve a determinate equilibrium relationship
among the transactors and their incomes; and (4) the allocation of
income described is an allocation among functional classes.  An addi-
tional assumption, that of relatively unrestricted competition (includ-

9. Lionel Robbins, *op. cit.*, p. 67; see entire discussion, pp. 63–71.
1. E.g. Joan Robinson's *The Economics of Imperfect Competition*, Chaps.
XXI to XXVI deal with labor, while J. R. Hicks has dealt separately with capital
and labor in *Value and Capital* and *The Theory of Wages*.

ing the hypothesis of a multitude of sellers and buyers, each controlling a small fraction of the total demand and supply, and the hypothesis of high geographical and functional mobility of resources and products), was characteristic of maximization theory before it began the assimilation of the theories of limited competition (i.e., monopolistic and imperfect competition), and is still accepted as in the main true by a large proportion of contemporary maximization economists. Clearly, then the problem which the maximization economists answer is not the problem which most of them appear to ask. This divergence is not *per se* a fault; it represents a type of abstraction often required in theoretical inquiry. By narrowing the problem as it has done, the school has provided the most elaborate, penetrating, and systematic analysis of the ultimate problem to the present date. Moreover, the number of its members who have been prepared to revise their working problem has been sufficiently large so that, in the course of its evolution the school has swallowed many dissidents and heretics to its own advantage. Perhaps it will continue to do so in the future, as it should, if it lived up to the claims of those of its representatives who identify it with the entirety of economic science. But with each absorption, its working problem and its doctrinal system undergo change.

On the other hand, those maximization or orthodox economists who today identify the maximization equilibrium problem with *the* problem (i.e., the ultimate problem) of distribution theory are guilty of a type of Philistinism ill-becoming the practitioners of a science. By dismissing or ignoring the members of other schools of distribution theory as noneconomists or as bad economists, because they fail to provide an answer to the maximizationists' operational problem, these maximizationists would, and to a degree actually do, break off communication between themselves and the remainder of the scientific world, congratulate themselves on having won a race from which they have excluded all competitors, and stultify the development both of their own school and of economic theory.

The second group of economists to be examined here falls within what might be called the "property school." This school includes all economists who attempt to explain the distribution of income in the private market system mainly in terms of the institution of private property. Although the school, so defined, is subsumed almost entirely within the "historical school," as defined below, it is important enough to merit separate treatment.[2] The property school is ancient

2. Not quite identical with the property school is the "socialist school," or more broadly the "collectivist school," the latter being defined to include all

in lineage (in modern form it is quite as old as the maximization school) and manifold in variety, but its branches have become so completely overshadowed in modern times by the branch known as Marxism that our attention will be limited here to the "Marxist school." Like the maximization school, the Marxists are at least comprehensive enough to constitute a school of economic, and not merely of distribution, theory, but will be examined here specifically in the latter aspect.

A difficulty arises in the selection of authoritative present day spokesmen for the Marxist school, because the latter, like the property school of which it is a part, has split into numerous factions. Since the differences among these factions are almost entirely outside the problem of capitalist distribution theory it might seem appropriate to quote from the group currently enjoying greatest power and prestige, commonly known as the "Marxist-Leninist-Stalinists," or "Stalinists" for short. This group constitutes a school in the subjective as well as objective sense defined above (p. 443); it is at core a monolithic organization with common principles subscribed to by all members and accepted by their peripheral sympathizers. This seems to simplify our task: the citation of the views of any member would appear to be representative at any moment of the views of all of them. However, the further away an economist is from the hub of the school (and it is not always easy for him, or for us, to know how close he really is) and the greater the time interval between his assertion and our citation of him, the more risk we assume in regarding him as an authoritative spokesman. The permanent or temporary fall from grace of authoritative economists of Stalinism in the past (e.g., N. A. Voznesensky, Eugene Varga, Karl Radek, or Nikolai Bukharin) indicates the hazard of quoting any current representatives or sympathizers, such as K. Ostrovityanov, A. Leontiev, or, closer home,

persons (i.e., communists, socialists, syndicalists, etc.) advocating an economy in which the private market sector would be secondary or nonexistent and in which, to a major extent, the natural resources, capital, or both would be owned, or the allocation of resources and incomes would be controlled, by society — organized either as a state (democratic or dictatorial), or as a group of associations (cooperatives, industrial unions, guilds), or as a brotherhood of highly moral individuals (anarchism). The property and collectivist schools differ in that some economists who emphasize the effects of private property on distribution do not definitely advocate a basic change in this institution in the direction mentioned (e.g., J. C. L. S. de Sismondi or Rudolf Stolzmann) while others who favor such a change accept a theory of distribution in which private property plays no important role (exemplified perhaps by some of the Fabian Socialists and Veblenians). Inasmuch as the schools are being distinguished in this paper according to their methodology in describing the private market system rather than their ethical or pragmatic objectives, the "property school" concept is here preferred.

Maurice Dobb or Paul Sweezy. It will be safest therefore to follow the practice of cautious Marxists of quoting directly from the writings of their "founding fathers."[3]

To Marx and his collaborator, Friedrich Engels, there are at least two distinct problems of distribution theory. The most important one, which Marx called "production of surplus value," might be approximated for our purposes as "distribution in the aggregate," i.e., the apportionment of income between the capitalist class as a whole and the working class as a whole. The other, which Marx regarded as a problem of "circulation," might be called "distribution in detail," and refers to the apportionment of the income of the capitalist class among its subclasses (capitalists proper and landlords) and among their divisions down to the individual capitalists themselves. Marx discusses the first problem most fully in the first volume of *Capital*, the second partly in the second volume, and both problems combined in the third volume.

The working problem of distribution theory to Marx and Engels, embracing the two problems, consists in the search for the laws of "exploitation" or "surplus value" in the capitalist economy (though the latter phrase sometimes appears restricted to the aggregate problem alone). A clear indication of this is the title of Marx's history of the distribution theories of his predecessors, *Theorien über den Mehrwert*,[4] which Engels describes as "an exhaustive critical history of the main point of political economy, the theory of surplus value. . . ."[5]

Some of the main conditions for the solution of this problem are mentioned by the authors in several places. Thus Marx's *magnum opus* introduces the problem of the production of surplus value as follows: "The conversion of money into capital has to be explained on the basis of the laws that regulate the exchange of commodities, in such a way that the starting point is the exchange of equivalents. Our friend, Moneybags, who is yet only an embryo capitalist, must buy his commodities at their value, must sell them at their value, and yet at the end of the process must withdraw more value from

3. This unavoidable identification of current Marxism with the writings of Marx, it has often been pointed out, probably would be resented by the great economist who, as his personal friends and contemporaries attest, "did not wish to be called a 'Marxist' and ridiculed the 'Marxists' to his heart's content." (Wilhelm Liebknecht, *Karl Marx: Biographical Memoirs*, transl. by E. Untermann, Chicago: Charles H. Kerr & Co., 1901, p. 71.)

4. Edited by Karl Kautsky, 4 vols. (Stuttgart: J. H. W. Dietz Nachf., 1905–10).

5. Preface to *Capital*, Vol. II, p. 8. This and the following citations from *Capital* refer to the translation by S. Moore, E. Aveling, and E. Untermann, 3 Vols., (Chicago: Charles H. Kerr & Co., 1906–9).

circulation than he threw into it at starting. His development into a full-grown capitalist must take place, both within the sphere of circulation and without it. These are the conditions of the problem. Hic Rhodus, hic salta!"[6]

At the end of the ensuing analysis of the problem in its general form, Marx concludes: "Every condition of the problem is satisfied, while the laws that regulate the exchange of commodities, have been in no way violated. Equivalent has been exchanged for equivalent. For the capitalist as buyer paid for each commodity, for the cotton, the spindle, and the labour-power, its full value. . . . He sells his yarn at eighteenpence a pound, which is its exact value. Yet for all that he withdraws 3 shillings more from circulation than he originally threw into it. This metamorphosis . . . takes place both within the sphere of circulation and also outside it. . . . "[7] Thus the problem in its introductory, general form is solved.

In a different connection, the problem is restated by Engels: "Whence comes this surplus value? It cannot come either from the buyer buying the commodities under their value, or from the seller selling them above their value. . . Nor can it come from cheating. . ." "This problem must be solved, and it must be solved in a *purely economic* way, excluding all cheating or the intervention of any force — the problem being, how is it possible constantly to sell dearer than one has bought, even on the hypothesis that equal values are always exchanged against equal values?" "The solution of this problem was the most epoch-making achievement of Marx's work."[8]

From these citations it is evident that the working problem of distribution theory to Marx and the Marxists is to explain how the owners of nonproductive factors (capital and land) receive a surplus (profits, interest, and rent) above their investment (i.e., above the cost of production of capital) even though the remaining factor (labor) is sold at its value and its products taken in the aggregate are sold at their value. Thus the operating problem of the Marxists differs from the ultimate problem (p. 444) by the major working assumptions that (a) certain environmental conditions remain fixed, including, e.g., the "relations of production," (b) the transactors maximize their net incomes under the given conditions, (c) only one social class or factor (labor) is productive beyond its own replacement cost, (d) a separate social class (capital) owns the means of production, (e) competition

6. *Ibid.*, Vol. I, pp. 184–5. See footnote also.
7. *Ibid.*, p. 217.
8. *Herr Eugen Duehring's Revolution in Science* (1878), transl. by E. Burns (New York: International Publishers, 1935), pp. 231–2.

prevails, on the whole, within each of the two classes, (f) all commodities, at least in the aggregate or on the average, exchange at their values (in the sphere of "circulation"), (g) value is equal to cost, in the form of labor time socially necessary for reproduction of the commodity, (h) despite exchange at value, the capitalists withdraw a net income.

The Marxist working problem has much in common with that of the maximizationists, but with important differences. (1) Although both assume an environmental *ceteris paribus*, their environments are not identical; particularly, the Marxists incorporate as an integral part of their subject matter the development of technology in response to and in reaction upon other features of the economic organization (an application of the materialistic conception of history), emphasizing especially its influence upon the organic composition of capital. (2) Although both assume economic incentives, rationality, and knowledge of market conditions to be prevalent among transactors (i.e., the maximization principle), the Marxists adopt this hypothesis only with respect to transactor choices in the private market system; they regard rational action by workers to consist more fundamentally in organization into trade unions — initially to prevent wages from falling below the subsistence level; eventually, to overthrow the private market system itself. (3) Although the essence of Marx's theories of aggregate and detail distribution can be expressed without assuming competition, the particular form of his answer, especially insofar as it includes a theory of value, prices, production, and of a uniform rate of profit and interest, does in fact assume (like the maximizationists prior to the incorporation of the limited competition theories) the general operation of unrestricted competition. The one exception is an assumed functional mobility of almost zero from labor to capital — two noncompeting groups. The fact that Marx infers from the competition within each class a necessary tendency toward monopoly does not remove the above assumption from his distribution theory any more than his assumption of a private market is removed by his prediction of its demise.

The most important differences between the Marxist and maximizationist assumptions are their concepts of productivity and of equilibrium. (4) Marx's notion that labor alone is productive has much in common with the views of the classical economists but is opposed by the maximization school as a whole, especially in its postclassical forms. (5) Marx postulates long run equilibrium values (or prices of production) and incomes, around which market prices and incomes fluctuate; but Marx's equilibrium is a dynamic rather than a

static one (even though the environment is assumed constant) in that certain recurring disequilibria are of such generality, amplitude, and duration as to constitute "crises" or business cycles while a basic contradiction or instability in the system leads to a declining secular trend eventuating in capitalism's self-destruction.

Like the maximizationists, then, the Marxists define the operational problem of distribution theory in a form considerably narrower than the ultimate problem common to both of them. Marx's problem, historically viewed, was not the abstract problem of explaining distribution, nor even exploitation, in capitalism but of proving that the hypothesis of capitalist exploitation was compatible with the essential features of the economic science of his time. For this purpose, proof that exploitation exists or that Ricardian economics is firmly grounded are equally unnecessary; Marx's major efforts at proof are to be found not in his agreements but in his differences with utopian socialists and classical economists. This explains why some of his doctrines such as the labor theory of value, which in terms of present day viewpoints look like conclusions, actually are assumptions and have been classified above as such. Marx succeeded so admirably in solving his problem that he was largely instrumental in forcing the maximizationists to replace some of the classical hypotheses with what they regard as more adequate substitutes.[9] Present day Marxists are still engaged in proving that capitalist exploitation is compatible with the assumptions of the classical school, the school which Marx helped to bury a century ago. Contemporary economists who do not participate in answering the Marxists' problem because they do not accept its exploitational or its Ricardian assumptions, or both, are condemned as "capitalist lackeys," "tools of Wall Street," or "apologists of wage slavery." But the Marxists do not thereby transform their assumptions into well-substantiated conclusions nor their working problem into the ultimate problem.

The combination by Marx of the exploitational and Ricardian assumptions into his working problem of distribution left little room on essential points for a solution other than his own. This may be shown by considering briefly some alternatives to his answer. One alternative solution, that capital and land as well as labor produce more than their cost is quite effectively ruled out by Marx's assumption that labor alone is the "value-creating substance." A second, that future products are worth less than present products, the difference constituting surplus value (Boehm-Bawerk's agio theory) is ruled out by Marx's assumption that homogeneous labor alone, in the

9. For a similar opinion see quotation from J. M. Clark on p. 467 below.

form of socially necessary units of simple average labor time, is the ultimate measure of value; present and future goods would differ in value, according to Marx, only in proportion to any change in the labor time necessary to produce the future goods and not (as in Boehm-Bawerk) in proportion to the time interval elapsing between the present and future goods and the degree of general preference for present over future goods. A third alternative, that surplus value originates in the sale of products (e.g., Wilhelm Lexis' theory), is ruled out by the assumption that all commodities exchange at their value.

Parenthetically, one may argue that Marx does not follow the conditions of the problem postulated by him, e.g., that all commodities exchange at their value, for capital in his system sells above its value (labor time socially necessary for its reproduction) as does land.[1] If the Marxists answer that this postulate applies only to the hire of labor and the sale of products, or that land and capital are not "commodities," or that their prices are "irrational" (Marx's term), then it becomes even clearer that the problem is so stated as to imply Marx's solution. To a lesser degree Marx violates the same postulate by substituting "prices of production" for "values."[2] Here he could reply that the sum or average of all products sells at the sum or average of all their values. But his postulate is violated in proportion to the dispersion around the averages — i.e., the variations in organic composition of capital.

The Marxists, like the maximizationists, commit the error of asking an answer-begging question, thereby guaranteeing that their positive doctrine will turn out to be a question-begging or more specifically an assumption-begging answer.

The working problem of distribution theory for the historical school is more difficult to ascertain than those of the two preceding schools. By the historical school is meant those economists who, on the whole, reject the methods and doctrines of the other traditional schools, mainly on empirical grounds; who usually are, or think they ought to be, engaged primarily in empirical research (statistical or otherwise), which they regard as a more valuable contribution to economics than long chains of deductive reasoning; who, in the main, emphasize the importance of diversity and of change in the economic subject matter and therefore the relativity of economic descriptions; who sometimes have attempted to replace the static laws of the maximizationists with laws of historical development; who, whether

1. *Capital*, Vol. III, pp. 415–20, 448–9, 730–32, 759.
2. *Ibid.*, Vol. III, Chap. IX.

or not they reject the traditional distinction between economics and the other social disciplines, explain economic phenomena and their changes partly in terms of social, political, or legal influences; and who, finally, stress the bearing of these socio-historical influences upon the economic views not only of laymen but of economists themselves. Some of these tenets are more in the nature of an ideal program[3] than a realized achievement.[4] The school's philosophy has been well epitomized by one of its early members, Richard Jones:" If we wish to make ourselves acquainted with the economy and arrangements by which the different nations of the earth produce or distribute their revenues, I really know but one way to attain our object, and that is to look and see."[5] Among the most frequently cited representatives of the historical school may be mentioned Friedrich List, Wilhelm Roscher, Bruno Hildebrand, Karl Knies, Gustav Schmoller, Lujo Brentano, Georg F. Knapp, Karl Buecher, Werner Sombart, and Max Weber of Germany; Richard Jones, Thorold Rogers, Thomas Tooke and William Newmarch, Walter Bagehot, T. E. Cliffe-Leslie, Arnold Toynbee, W. R. Ashley, William Cunningham, John K. Ingram, and Sidney and Beatrice Webb of England; J. C. L. S. de Sismondi, Georges d'Avenel, Emile Levasseur, and Henri Sée of France; and Thorstein Veblen, John R. Commons, Robert Hoxie, Walton Hamilton, Wesley C. Mitchell, and Gardiner C. Means of the United States, all of whom (despite important differences) have been called at some time or another "institutionalists." If such writers as William Petty or the mercantilists, Colbertists and cameralists are classified as historical economists, this school is indeed of early origin.

Numerous differences exist among the various branches of the school, which cannot be considered here. Emphasis will be on their common features, though the institutionalists will be used as the example closest at hand. The term "sociological school" may be used synonymously with "historical school."

Historicism should be distinguished from historical literature. Some of the best contributions to historical literature have been made, of course, by economists equally well or better classified into other

3. E.g., Wilhelm Roscher's *Grundriss zu Vorlesungen ueber die Staatswirtschaft* (Goettingen: Dieterichsche Buchandlung, 1847) or various essays in Rexford G. Tugwell (ed.), *The Trend of Economics* (New York: Alfred A. Knopf, 1924). The latter often substitutes "will" for "should."

4. A keen interest in social reform also characterizes many historical economists, but this is irrelevant to the present essay.

5. "An Introductory Lecture on Political Economy" (1833), in his *Literary Remains* (London: John Murray, 1859), pp. 568–9.

schools.[6] And some of the most important writings of the major historicists have been little more than economic histories or statistical compilations, which are not necessarily inconsistent with the theoretical doctrines of other schools. To qualify as a historicist, however, it is more necessary that one have the empirical viewpoint than that one contribute to empirical research. For this reason, although all empirical surveys may be regarded as part of historical literature, only those economists who have expressed a pro-empirical theory in some part of their writing, even if it is no more than a brief article or the preface to an empirical study, will be regarded here as historicists.

The historical school, like the two others, covers the entire range of economic theory, but is here examined primarily in relation to the distribution problem. An examination of the writings of historical economists reveals a paucity of efforts, however, not only in defining but also in dealing with the distribution problem as a general and distinct problem of analysis. The reasons for this may be found in the nature of the historical viewpoint and include the following:

First, most historicists have an extreme distrust of theoretical reasoning; they are more interested in extrovert fact finding than in methodological self-consciousness and are impatient to "look and see" rather than to spend time in preliminary clarification of what they are to look for and how they are to look for it.

Second, even to the extent that they clarify their problems, the problems to which their empirical bent inclines them are considerably less abstract than that of "capitalist distribution." It is the multitudinous varieties and evolutionary changes in capitalism and in capitalist distribution rather than the features shared by these phenomena in common or within capitalism's private market sector which historicists regard as meaningful subjects of investigation. Without a set of special lenses it is easier to look at, and to find data on, wages in the automobile industry in Detroit in 1949 than to examine wages in capitalism.

Third, to the extent that the historicists deal with general prob-

6. E.g., Adam Smith on the mercantile system in the *Wealth of Nations* and on social mores in *The Theory of Moral Sentiments*, T. R. Malthus on the standards of living of various nations in the second edition of his *Principle of Population*, J. S. Mill on land ownership systems in his *Principles of Political Economy*, W. S. Jevons on *The State in Relation to Labour*, Alfred Marshall on *Industry and Trade*, R. T. Ely on *Property and Contract in Relation to the Distribution of Wealth*, or F. W. Fetter on *The Masquerade of Monopoly;* and J. C. L. S. de Sismondi on business cycles, Friedrich Engels on *The Condition of the Working Class in Great Britain in 1844*, or Karl Marx on "the myth of primitive accumulation" and on the struggle for the ten-hour day in Volume I of *Capital*.

lems, their subject matter usually consists in what are called "economic institutions," which cut across the problems of the other traditional schools.[7]

Fourth, even to the extent that the traditional divisions are recognized, many historicists regard the traditional problems of value and distribution as relatively unimportant. Thus Wesley C. Mitchell predicted, with implied approval: "Another change that will come over economics is the recession of the theory of value and distribution from the central position it has held ever since the days of Ricardo to make room for the theory of production."[8]

Fifth, historical writers of a theoretical bent are clearer in their criticisms of traditional theories than in their positive views. After demolishing the theoretical structures of their opponents, some of them, when they attempt to present their own doctrines, are so discursive, complex, incoherent, or vague that it is impossible to summarize their positions. Others, when they seem on the verge of presenting a clear definition, a general uniformity, or a "realistic" theory, will suddenly relapse into one of several devices which leave the reader in the same negative state he was in when the "unrealistic" edifices were demolished. One such device is to "define by example" — to substitute a particular datum for a general concept. Another is to "define by contrast" with the wrong theory, thereby telling us once more what the correct theory is not, rather than what it is. When confronted with the problem of explaining some specific phenomenon (e.g., the wage rates of automobile workers in Detroit in 1949) a favorite device is that of presenting a "historical background," which may be long or short in chronology, wide or narrow in geography, and remote or direct in relevance; which often assumes that *ante hoc ergo creandum hoc;* and which usually neglects to show how the historical background as described necessarily produced the phenomenon in

7. Thus one may cite Brentano on *Der Arbeitergilden der Gegenwart;* or Sombart on *Der Bourgeois, Luxus und Kapitalismus,* and *The Jews and Modern Capitalism;* or Weber on *Wirtschaft und Gesellschaft* and *The Protestant Ethic and the Spirit of Capitalism;* or Veblen on *The Theory of the Leisure Class, The Theory of Business Enterprise, Absentee Ownership,* and *The Instinct of Workmanship;* or Commons on *The Legal Foundations of Capitalism, The Principles of Labor Legislation,* and *Institutional Economics;* or Hamilton and May on *The Control of Wages;* or Berle and Means on *The Modern Corporation and Private Property;* or Wesley C. Mitchell on *Business Cycles.* The authors of these studies are not specifically concerned with the effects of the institutions or laws examined upon value, prices, wages, rent, interest, and profits; their concern may be considerably narrower or broader.

8. "The Prospects of Economics," in Rexford G. Tugwell (ed.), *The Trend of Economics* (New York: Alfred A. Knopf, 1924), p. 29. See also articles by Walton H. Hamilton, cited in footnote 4 on p. 441 above.

question rather than some other phenomenon. Such theoretical deficiencies are not surprising in economists who are suspicious of ratiocinative processes and feel more at home in empirical research.

Finally, those economists allegedly members of the historical school who do present a relatively clear, positive system of doctrines generally turn out to belong (at least as appropriately) to some other school of distribution theory. Some, like the later Roscher or John Neville Keynes, are quite as much maximizationists as historicists. Others, like the Webbs, J. A. Hobson, Franz Oppenheimer, or H. D. Dickinson, fall as readily into the property school.[9]

What, then, may we conclude, first, as to the nature of the ultimate problem of distribution as viewed by the historical economists? It is legitimate for us to depict this problem, from the historicists' view, as being identical with that of the other schools. (a) The fact that some historicists, in practice at least, indicate an unwillingness to accept the legitimacy or importance of the distinction between the problem of distribution and other problems (e.g., production) constitutes no serious difficulty here. The distinction mentioned accords with pragmatic issues and popular usage, it is accepted by most economists as having heuristic value, and its acceptance does not negate the possibility that there may be other fruitful modes of dividing economic phenomena — e.g., by type of institution. (b) The fact that many historicists do not regard the distribution problem as important leads, at most, to the inference that the present essay is unimportant. It has often been pointed out, however, that the fact that market transactors themselves regard the question as important (e.g., collective bargaining disputes or the attitude called "keeping up with the Joneses") does make the question important and from the historicists' own "behavior" viewpoint. (c) Some members of this school may object to using the term "distribution theory" in the statement of the ultimate problem and wish to substitute "study of distribution" instead. This change is entirely acceptable here, although the term theory as used above is intended to include empirical as well as deduced generalizations. (d) Finally, some may wish to emphasize that there is not a single distribution problem with a single general solution but a host of small distribution problems requiring a host of discrete analyses. This change also is acceptable, although the term problem as used above is intended to include the

9. The latter group of economists could also be classified, with Schmoller, Rudolf Stolzmann, Karl Diehl, Commons, or Paul Schroeder, into a fourth school, which may be called the "power school." This is not one of the traditionally recognized major schools and therefore is not considered in the present essay.

entire field within which numerous problems can be found, and is not intended to prejudge a monistic solution. The ultimate problem of distribution theory to the historical economists, then, may be regarded as identical with that of the maximizationists and Marxists.

But what can we say of the working problem of the historical school? From some of the statements on economic methodology by historicists, especially their strongly expressed antipathy toward *a priorism,* it is easy to get the impression that their working problem is identical with the ultimate problem. This unqualified identification is tenable, however, less as a description of their actual behavior than as a description of the description which some of them apply to their behavior. True, their suspicion of deductive theorizing and categorizing leads them to avoid the conscious narrowing of their inquiry by the conscious adoption (in advance of the major part of their fact-gathering activities) of a set of straitjacketing assumptions. To the extent that they succeeded in following this *tabula rasa* approach, they would be left to grapple with the huge, formless, Protean mass constituting the economic subject matter, equipped, at best, with bare hands, native intelligence, and will power. Fortunately, they do make assumptions or continue to entertain assumptions acquired during their previous education and experience, exactly like the theorists whom they disparage. But to the extent that these assumptions are made unconsciously (on the further assumption that they are making no assumptions) the extreme historicists increase their probability of error relatively to that of self-conscious theorists.

In practice, then, historical economists like any others attack their problems armed with assumptions, which tend to limit those problems themselves. Thereafter, they usually proceed to follow one of three courses of action. One is to break off a piece of the subject matter, small in space, duration, and other dimensions; examine it minutely; and conclude that it is very difficult, for a number of reasons to generalize from it. A second is to compile statistical and other data concerning somewhat larger areas of the subject matter, e.g., time series on prices, interest rates, or wages, which are useful in proportion as they can test pre-existing hypotheses, generally those advanced by the members of other schools, and thus lead to less objectionable new ones. And a third is to discourse theoretically, unsystematically — and at times brilliantly — on certain alleged social uniformities of broad scope, touching occasionally upon their effects on income distribution.

Whether the working and ultimate problems of historicists are identical reduces itself then to the question whether the uniformities

which they seek by these three procedures are in any way more con-
crete than uniformity in the abstract which all scientists in the field
are bound to seek anyway. One method of achieving this concrete-
ness is to select only certain types of uniformities and ignore others.
A second is to concern one's self with all or most uniformities but to
divide them into distinct, clearly defined types and work on each
somewhat independently. Both methods, while introducing con-
creteness into the problem and providing a criterion as to procedure
in handling it, constitute, be it noted, an *a priori* delimitation of the
scientist's procedure from the time of their adoption onward.

Most historical economists follow either or both practices to some
extent, especially the American institutionalists, in so far as they center
their analyses around institutions and their attendant laws, customs,
conventions, traditions, folkways, mores, and the like. But the actual
degree to which they channel their working problem definitely and
clearly varies greatly from one economist to another. The term insti-
tution is highly ambiguous and institutionalists are rather wary of
defining it, as they are anything else. They have described institu-
tions variously as "settled habits of thought common to the generality
of men" (Veblen), "collective action in control of individual action"
(Commons), "segments of social behavior predominantly ceremonial
in character" (Clarence E. Ayres), etc. Some have identified as insti-
tutions the family, state, church, industrial technology, private prop-
erty, free contract, business enterprise, political democracy, national-
ism, trade unions, business associations, market structures and agree-
ments, absentee ownership, advertising, the keeping of accounts, con-
spicuous consumption, monetary system, imperialism, business cycles,
and membership in the Elks or Masons. Of these social uniformities
it will be noted that some result from imitation of others past and
present, some from coercion by the state, some from pressure by large
organizations on non-members, some from discipline imposed by
organizations on their own members. Business cycles occasionally
are referred to as institutions, though whatever uniformities they
possess are no more habitual or voluntaristic in origin than are attacks
of hay fever. Thus some historicists may include under institutions
practically all types of social uniformities except those resulting from
natural constants, from acts of rational choice, from mutual interac-
tion of conflicting individuals (e.g., orthodox laws of equilibrium), and
from chance.

Yet many historicists regard "constants in human nature" as
either very restricted in number, or as institutions unrecognized as
such by those who follow them, or as propensities which materialize in

institutions or which can be prevented from materializing by other institutions (e.g., Veblen's "instinct of workmanship"). Many regard rationality as quite uncharacteristic of the human species or as an "enlightened common-sense" which itself is rooted in institutions of time and place. The theory of an unplanned balancing of opposed forces, such as the orthodox laws of static equilibrium, historicists dismiss at best as an abstraction of certain tendencies uncharacteristic of the dynamic, nonteleological processes in the real economy. And finally, many historicists, especially of the statistical branch, broaden their subject matter to include all uniformities of economic life capable of quantitative measurement, whether or not the uniformities are called institutions and regardless of the source of the uniformities.

One may conclude, then, that the historicists' working problem of distribution consists in the study of (a) social uniformities of a quasi-voluntaristic nature (institutions, laws, and customs), (b) social uniformities resulting from voluntary acts of large social groups (e.g., wars, revolutions, collective bargaining, monopolistic agreements), and (c) other social uniformities which can be expressed at least partly in statistical terms (e.g., business cycle series, such as rigid and flexible prices), in so far as these three types of uniformities and their changes and varieties throw light upon the apportionment of the total product in a private market system. The degree to which this working problem, in contrast with the ultimate problem, achieves concreteness without a corresponding loss of generality of scope, varies considerably from one historicist to another, depending on his breadth of interest, his ability to distinguish the various types of social phenomena, and his ability to construct a general framework for his system of ideas.

That some historicists do possess a working problem and working assumptions of some concreteness, like the members of other schools, may be shown briefly by citing the issue of personal or institutional versus functional distribution, which pervades much of the historical literature, though the terms mentioned are not always employed. Historicists have insisted that a given division of incomes cannot be explained more than superficially by reference to the functional categories of labor, capital, land, and enterprise; that one must look behind these categories for the social influences which determine why some individuals and not others possess capital or land, or are able to act as enterprisers, as practitioners of professions, or as skilled workmen. For this purpose one must examine the institutions and

laws of inheritance, taxation, education, and the like.[1] The maximizationists' main reply, of course, is that the historicists have rambled outside the borders of "economic causes." But the issue is fairly concrete and implies some concreteness of assumption on both sides. It cannot be denied, however, that the historicists on the whole have not proceeded very far in the construction of a systematic body of working assumptions and consequently of a systematic body of doctrine. Their working problem is still disappointingly close to the ultimate problem, giving them much freedom but also much bewilderment as to what to do with it.

Finally, the historical economists have their own peculiar ways of dismissing the theories of their opponents, one of the most effective (borrowed from the Marxists) being to regard these theories historically as "ideologies," expressing the institutionally molded desires of the social classes or groups to which the economists in question belong, and reshaping, in turn, the customary beliefs and actions of the community.[2] The theories of maximizationists and Marxists are thereby relegated to the sphere of causes and effects constituting the subject matter of economic study; thus there is no need to consider them as rival descriptions of that subject matter.

## V

The terms "isolationism," "exclusivism," or "narrowness," used here as roughly synonymous, are not intended to apply in equal degree to each school as a whole or to its members. Isolationism may be analyzed into at least four distinct components, with respect to (a) the range of phenomena examined, (b) the range of methods used, (c) the frequency of communication with other schools, and (d) the tenacity with which assumptions concerning the first two, and conclusions derived from them, are held. Estimates of isolationism are extremely difficult to make in quantitative terms, particularly for a school as a whole and in all four respects combined. However, several observations can be made which appear capable of substantiation. With respect to the range of phenomena, more specifically, the range of

1. The controversy between the Marxists and maximizationists over "primitive accumulation" is a special subdivision of the same issue.
2. "The notion of a hypothetical equilibrium which would result if certain theoretical requirements were fulfilled is introduced [in this book] only in a limited way; and not because it is a description of tendencies actually operating in the world, nor because the authors believe that orthodox theory provides an adequate tool of analysis, but because the ideal of equilibrium has had a profound effect upon economic behavior." (Donald W. McConnell and Others, *Economic Behavior*, revised ed., Boston: Houghton Mifflin Co., 1939, p.vi.)

causal influences on market distribution, the historical school appears to be broader than the Marxian school and the latter broader than the maximization school. The frequent accusation by the maximization school that the historical economists and to a lesser degree the Marxists are largely "sociologists" is an indication of this fact.[3] With respect to the variety of methods actually used, there is no preponderant difference visible among the schools. Maximization and property school members have contributed some of the best empirical studies to the science (p. 459, note 6, above). On the other hand, one finds deductive reasoning developed to a fine degree in the writings of Max Weber, Thorstein Veblen, Arthur F. Burns and Wesley C. Mitchell.[4] Karl Marx used theoretical (including mathematical) and historical (including statistical) techniques while among contemporary Marxists are to be found theoreticians like Maurice Dobb and empiricists like Juergen Kuczynski.

With respect to intercommunication, the three schools have always been aware of one another at their fringes; bold members of each have repeatedly sallied forth to inspect, learn from, or demolish the outlanders and at times open warfare has broken out among them. Thus a sizable "exogenous" literature has accumulated, in which may be included the *Methodenstreit* led by Menger and Schmoller; the "original accumulation" issue; the controversies among the Marxists, Austrians, Fabians, and Revisionists from about 1890 to about 1920; and the running disputes on economic psychology, control, rigid prices, and the like between the institutionalists and the maximizationists in the United States from about 1900 onward.[5] But this critical literature is not quite the same thing as intercommunication.

3. One of the most persistent as well as sophisticated of those distinguishing between "economics" and "sociology" along the lines indicated is Joseph Schumpeter. See, e.g., his "Das Grundprinzip der Verteilungstheorie," *Archiv für Sozialwissenschaft und Sozialpolitik*, XLII (1916–7), 16–26 and his "The Communist Manifesto in Sociology and Economics," *Journal of Political Economy*, LVII (1949), 203–4.

4. See, e.g., the last two authors' *Measuring Business Cycles* (New York: National Bureau of Economic Research, 1946).

5. Of the handful of recent books largely of an exogenous nature in the field of distribution theory there may be mentioned those of the maximizationist Joan Robinson, *An Essay on Marxian Economics* (London: Macmillan and Co., 1942); the Marxists, Maurice Dobb, *Political Economy and Capitalism* (2d ed.; New York: International Publishers, 1945) and Erich Roll, *A History of Economic Thought* (New York: Prentice-Hall, Inc., 1942); and the historicists Clarence E. Ayres, *The Theory of Economic Progress* (Chapel Hill: University of North Carolina Press, 1944) and Richard A. Lester and Joseph Shister (eds.), *Insights into Labor Issues* (New York: The Macmillan Co., 1948). Some of the authors in the latter study, however, appear to be basically maximizationists, criticizing from within. The periodical literature is too large to cite here.

To a large extent it is written by the members of a school (usually a minority school) for perusal by other members of the same school, and consists largely in the demonstration that the working assumptions of another school (usually the dominant school) are inconsistent with their own. "Isolationism" with respect to communication consists more in unwillingness to receive than to give criticism, and as such is much more difficult to estimate.

All three schools have shown at times extreme intolerance toward criticisms by others, although the present day maximizationists and historicists, at least in the United States and England, are far more receptive than the Marxists, whether inside or outside the Soviet orbit. Nevertheless only a small proportion of the articles published by the western economic journals are critical of the school to which the editors and readers on the whole belong.

Finally, judging narrowness by obstinacy in adhering to doctrines is also a difficult task, especially since it begs the question whether the obstinacy persists in spite of the error or because of the truth of the doctrines held. In number of changes of doctrine during the past century, the maximization school probably takes the lead. Some of its changes have resulted from criticisms by the other schools, but it has not always acknowledged its debt. For example, J. M. Clark has observed: "The marginal theories of distribution were developed after Marx; their bearing on the doctrines of Marxian socialism is so striking as to suggest that the challenge of Marxism acted as a stimulus to the search for more satisfactory explanations."[6] Similarly, there is reason to believe that the development or acceptance of the imperfect and monopolistic competition theories was stimulated by the historicists' criticisms of the competitive assumptions of maximization theory prior to about 1930. A considerable gulf also divides the views of historicists of today from those of a century ago, or even from Veblen and *The Trend of Economics*, probably reflecting in part the benefit of maximizationist criticisms. Many present day institutionalists seem to be by temperament and by virtue of their own historical philosophy as open-minded as economic scientists should be. But their paucity of broad, positive generalizations and the gaps in continuity among the various branches of the school (e.g., the Schmollerites and Veblenians) make it difficult to speak of actual changes in view (as distinct from variation in subject matter) of the same group. No individual reversals of doctrine have appeared among the historicists comparable to, say, those of the maximizationists J. S. Mill or A. C. Pigou. The Marxists alone, admittedly, have learned nothing

6. "Distribution," *loc. cit.*, III, 170; *Readings*, pp. 64–5.

and forgotten nothing in the field of capitalist distribution theory.[7] They justify their tenacity, of course, on grounds of the fundamental accuracy of the Marxian system.

Despite some intercommunication and doctrinal change, the schools have succeeded in preserving their independent existence for over a century; if interpreted broadly, they go back approximately to the beginning of capitalism. The maximizationists have been the leading school during most of the past century in western Europe and to a lesser degree in the United States. The historicists have been dominant in Germany during approximately the same period. And Marxian economics has been the official doctrine in the Soviet Union for the past thirty years. When the schools have yielded to change, it has usually been on what its members and critics regard as unessentials. The changes seem to have been made not in the conviction that critics outside the school were right but that the doctrine thereby has gained so in elegance that it removes the ground beneath the critics' feet and makes further discourse with them unnecessary.

## VI

The origin of this exclusiveness is a difficult problem in the sociology and psychology of ideas, as well as in scientific methodology, and can only be touched on here. No implication has been intended above that the "oyster" tendency on the part of the schools arises from the same sources for all economists and all schools equally. The prolonged differences among the schools, like any other differences in ideas, may be explained in terms of influences that are either extraneous or integral to the field of inquiry and are either motives (conscious or unconscious) of the theorists or conditions faced by them. To what extent the insularity of the schools originates in extraneous motives, especially those of a political nature, is a particularly touchy issue, and may lead to the type of name-calling which the Marxists especially have made so familiar in their ideological warfare. The motives which contribute to the discovery or acceptance of an idea probably are extremely diverse, numerous, and complex in most cases, and never fully known to the economist himself. But no matter how foreign or germane to the field of inquiry nor how base or laudable the motives may be, they have no necessary bearing on the scientific adequacy of the idea. That the contemporary Marxists' motives in the arts and sciences are largely of a political nature is no secret; they

7. They have admittedly learned from bourgeois economists in other fields: e.g., Lenin's avowed acceptance of Hobson's thesis on imperialism. But even this was an addition, not a reversal.

are the first to admit this, in their proclamations of "proletarian art" and "proletarian science." That political influences affect the other economic schools is admitted without equivocation by some of their more self-conscious members, at least among the maximizationists.[8] Regardless of motives, the maximizationists' definition of economic science and its problems in such a fashion as to exclude as sociology the question of the influence of property and other institutions on economic phenomena diminishes the tendency of economists to blame those institutions for any allegedly undesirable economic phenomena and to adopt the notion that the institutions are natural and incapable of improvement by human intervention. In other words, the acceptance of institutions as inviolate "data" in economic methodology tends to encourage their acceptance as inviolate "data" in social praxis. If an economist did blame institutions or recommend changes therein he would have to do so, apparently, as an applied sociologist rather than as an applied economist. This statement, to repeat, refers to a result, not a purpose, of maximization doctrine. Conversely, historical and Marxian economics have the effect, apart from their respective motivations, of directing the economist's attention, while still an economist, to socio-political phenomena as possible explanations of undesirable economic phenomena. To some unknown but positive degree, these pragmatic consequences of the methodological positions of the three schools coincide with the conscious or unconscious political motives of their members.

Motives other than political, though still extraneous to the immediate problems of economic methodology, also have some influence on the methodology of economists. Inertia, fear of change, hero-worship, the pleasure of communication with kindred spirits rather than arguing with opponents, the desire to be on the winning side, the pecuniary impulse — these are just a few. The Marxists' conservatism in theory is required to some extent by their extremism in action. Economic self-interest undoubtedly influences some economists to follow the maximization school in the United States; the same motive, plus fear of personal safety, influences others to follow the Marxian school in the Soviet Union. On the other hand, sym-

8. Thus A. C. Pigou has said: "It is not wonder, but the social enthusiasm which revolts from the sordidness of mean streets and the joylessness of withered lives, that is the beginning of economic science." *Economics of Welfare* (4th ed., London: The Macmillan Co., 1932), p. 5. Or J. Schumpeter: "though we proceed slowly because of our ideologies, we might not proceed at all without them." ("Science and Ideology," *American Economic Review*, XXXIX [1949], 359.) See also Gunnar Myrdal, *Das politische Element in der national-oekonomischen Doktrinbildung* (Berlin: Junker und Duennhaupt, 1932).

pathy for the weaker side, or psychic abnormalities (the "Oedipus complex," "neurotic aggressiveness"), or the hope of gaining on the "wave of the future," or the wish to be different may induce economists to join the unpopular school. Once an economist has joined a school, his habits, desire to maintain face, loyalty to his comrades, etc., tend to keep him there. For mature economists a change of methodological *Weltanschauung*, in contrast with a change in doctrine, is almost equivalent to a change in personality. Indeed methodological differences such as those between the maximization and historical economists are in part "a conflict between two kinds of work, between men of different intellectual inclinations. . . ."[9]

But there are other sources of exclusivism which appear to be inherent in scientific activity. One consists simply in the desire to find the most efficient method for achieving scientific objectives. All economists to some degree have this motive, i.e., recognize the "rules of the game," no matter how strong their extraneous motives may be.[1] A second is the scientific desideratum of concreteness in problem and assumptions. On the one hand, the scientist strives to expand his working assumptions to meet the requirements of his ultimate problem. On the other, his working problem and assumptions must retain concreteness if he is to do any work at all. The method followed by some historicists in accomplishing the first of these consists in increasing the abstractness of their working assumptions (in the sense of expanding their breadth of coverage) by rejecting the method of abstraction (in the sense of selection of the important and removal of the unimportant). Claustrophobia here is cured by galloping off in all directions. On the contrary, as has been pointed out by maximization economists at least from J. S. Mill onward, simplified models are necessary in most, if not all, theoretical sciences.[2] Practising historicists themselves, it was shown above, have had to concretize their problem and method. This particularization, when done differently by different economists, contributes toward the insularity of the schools.

9. J. Schumpeter, "Epochen der Dogmen- und Methodengeschichte," in S. Altmann and Others, *Grundriss der Sozialoekonomik, I Abt.: Wirtschaft und Wissenschaft* (Tuebingen: Verlag von J. C. B. Mohr, 1914), p. 106.

1. The difference between "scientific" and "extraneous" motives is far from precise, especially since motives metaphysically presuppose others in an indefinite series. Ask a scientist *why* one method is more efficient than another, and you force him to appeal to criteria which you may then call "extraneous": e.g., the "aesthetic" criterion of "simplicity."

2. J. S. Mill, "On the Definition of Political Economy" (1836), in his *Essays on Some Unsettled Questions of Political Economy* (London: John W. Parker, 1844), pp. 138–41, 145–6.

A third methodological source of insularity is the isolation required by the scientist during his positive operations in constructing and refining his system of ideas and collaborating with empiricists or conducting empirical studies himself to test his hypotheses and conclusions. Most criticisms of his theories, he finds, come from colleagues whose structures closely parallel his own; his answers to them, therefore, facilitate rather than hamper his positive operations. On the other hand, to bother with external critics means to turn aside from building the edifice to securing the foundation, from conception to preconception, from a seemingly major to a seemingly subsidiary task.

A final methodological source of isolationism consists in the limits to the verifiability of propositions and assumptions in a complex and nonexperimental science like economics. A theory may be accepted or rejected in terms of whether it successfully runs the gauntlet of being (1) logically meaningful, rather than tautological or self-contradictory, (2) consistent with other accepted propositions, (3) empirically meaningful, i.e., expressed in terms making possible comparison with data, (4) practically capable of empirical test, i.e., whether the necessary data are, or can be readily made, available, and (5) actually tested by data. In none of these steps can one dispense with what is called theorizing or abstract reasoning, and an identical proposition may be declared defeated in every round by one economist while another equally competent economist at the end of the fifth round may judge it victorious. To take the fifth stage alone, our economist, whose theory appears to correspond with a series of data, may be told that the correspondence is accidental, due to changes in factors which he allegedly assumed to be constant; on the other hand, if his theory appears to be belied by the data he may answer that other things did not remain equal. Given the lack of controlled experiment, the data readily available are rarely in the proper form or adequate amount (especially in length of time series) to lead to reliable conclusions on issues which economists would call important. One assumption presupposes other assumptions in an indefinitely long chain; economic science may push back the borders of "meta-economics" by subjecting assumptions to empirical test, but an area of assumptions, not immediately verifiable, probably will always exist, providing room for opposing systems of theory which cannot for the moment be brought into genuine conflict or reconciliation.

## VII

The conclusions which this paper has attempted to establish may be summarized as follows: Each of the three traditional schools of

economic theory, though objectively a rival to the others in seeking a solution to identical ultimate problems of economic science, regards its members, on the whole, as the only genuine economic scientists and its operational problems and assumptions as the only legitimate problems and assumptions of economic science. Each school presents its doctrines not merely as the best but as the only available solutions not only to its own operational problems limited by its own operational assumptions but to the ultimate problems shared by all schools in common. Each school regards the other schools, to the extent that it regards them at all, from the viewpoint of its own operational problems and assumptions, and banishes them, therefore, from the territory of "economics," or of "science," or of both. Each school, although objectively a rival to the others, by envisaging its working problems as being identical with the ultimate problems, by substituting, that is, its own means for common ends, has established, in effect, a set of rules which disqualifies the other schools at the start and guarantees its own victory.

In the present essay these propositions have been examined in relation to only one problem (or set of problems) of economic theory — i.e., value and distribution. But this problem, interpreted broadly, is itself of sufficient breadth (if not also of sufficient importance) that the attitudes of the traditional schools toward it adequately represent their attitudes toward the problems of economic theory in general. The same conclusions here drawn concerning the three schools apply in much smaller degree to the attitudes of branches toward other branches within each school (the Marxian sects constituting an exception). Although the branches (e.g., monopolistic and imperfect competition within the maximization school) define their working assumptions differently from each other, communication among them is sufficiently close so that the working problem of each, in the sense of assumptions *examined* (not in the sense of assumptions *employed*) may be said to be practically identical with that of the others.

The exclusivist attitude, it is granted, varies considerably in degree from one school to another and from one economist to another. But a certain amount of it is clearly noticeable in all the traditional schools and in the majority of economists. This residue of exclusivism is in part a necessary good, in part a necessary evil, and in part an unnecessary evil.

It is a necessary good in the sense that it is required by scientific method, first (1) insofar as "charlatans," "rationalizers," "mystics," or "reformers" are excluded while "pure scientists" are included, or insofar as "sociologists," "political scientists," "psychologists," or

"historians" are excluded while "economists" are included. Nothing in the present essay is intended to belittle the importance of these distinctions. But any person who asserts that certain individuals are not economic scientists but belong in some other category distasteful to them must be prepared to shoulder the burden of proof in meeting the charge of slander. And he may not base his argument entirely on his own definitions of terms — quoting himself as his authority — unless he wishes to convince us all the more of his solipsistic attitude.

(2) Exclusivism is methodologically desirable, insofar as concreteness in problems and assumptions is desirable. But the scientific requirement of concreteness here means the need of some practicable limits to one's analysis; it does not imply simplification to any degree. On the contrary, in a broad and complicated field of inquiry one's working assumptions should be as numerous and complex as is practicable, i.e., as is consistent with other scientific requirements. One such requirement limiting the number of assumptions is Occam's razor. This, however, comes into play only when alternative chains of assumptions of unequal length lead to identical or equally defensible conclusions. A second limitation is the capacity of existing scientific techniques (and the human brain) in handling assumptions; the multiplication of assumptions beyond this capacity leads to intellectual indigestion. But simplified distortion is not necessarily preferable to incomprehensible accuracy. A dislike for both should lead to constant efforts to improve or replace existing scientific techniques to enable the manipulation of the more complex assumptions. The scientific techniques, moreover, do not necessarily consist exclusively in the techniques of the natural sciences nor in the ability to reduce phenomena to mathematical terms. And finally, the economist should not forget the restrictions on his working assumptions when he comes to his conclusions. It is not necessarily objectionable to postpone even indefinitely a revision of initially simplified postulates.[3] But it is unpardonable to state one's conclusions (or draw implications from them for pragmatic action) at the end of a theoretical study with an assurance greater than that with which the initial assumptions were presented or adopted.

Exclusivism is partly a necessary evil. (1) Thus far in the history of social sciences, the strongest motives to research appear to have been the "extraneous" ones of political reform, personal advancement,

3. Such as J. B. Clark's often quoted promise of a future study based on dynamic assumptions (*The Distribution of Wealth*, New York: The Macmillan Co., 1899, p. ix) or Alfred Marshall's modest reference to his theoretical treatise as "an introductory volume" (*Principles of Economics*, 8th ed., London, Macmillan and Co., Ltd., 1920, subtitle).

and the like, rather than the "indigenous" one of scientific curiosity. The same extraneous motives that are so potent in giving rise to new discoveries, unfortunately, also induce economists to hold on to out-worn ideas, to form cliques, to avoid facing new evidence. Partisanship is one of the costs of scientific progress. But it is a cost which itself retards scientific progress and should be reduced to a minimum.

(2) Exclusivism is a necessary evil insofar as it is bred by the isolation required for the construction of systems of ideas. But the marginal utility of positive construction in solitude eventually falls below that of testing the foundations by exposure to external attack or criticism.

(3) Exclusivism is a necessary evil to the extent that it originates in permanently or temporarily unverifiable assumptions and to the extent that such assumptions are unavoidable in economic inquiry. The persistence of meta-economics has been alluded to above. Autonomous experimentation by society is much slower in yielding significant data for use via the method of difference than is controlled experimentation in the laboratory. Conceivably, the returns of empirical data for the solution of a problem may come in so slowly that the subject matter itself may expire (to say nothing of the scientist) before the problem is ever solved. Unless further considerable improvements in methodology are made similar to the recent developments in statistical sampling techniques, highly arbitrary theorizing will continue to occupy a vast area of economic science. Within this area economists will continue to find room for the most diverse opinions without fear of disproof; insularity is thus protected by unverifiability. But this constitutes no justification for clinging to unverifiable assumptions when they can be replaced by verifiable ones. As the historicists especially have argued, it is highly doubtful whether such assumptions as "labor is the source or measure of all value" or "all transactors maximize their ends in relation to their means" possess, as they stand, empirical significance. The fact that certain unverifiable propositions cannot be replaced, for the time being, by verifiable ones, moreover, requires that they should be accepted not with the assurance that springs from their temporary impregnability to empirical attack but rather with the caution that befits their lack of substantiation.

Isolationism, finally, is to a degree a curable disease. (1) Negatively, it is a curable disease to the extent that, as has been mentioned immediately above, the methodological virtues of delimiting economic science and its problems and assumptions, and the necessary evils of extraneous motivation, solitary construction, and unverifiability of propositions, are carried beyond appropriate limits. The treatment

indicated here is the common acceptance of a line of division between the healthy operations in scientific method and their pathological excesses.

(2) Positively, isolationism is a curable disease to the extent that the excesses mentioned originate in the operation of extraneous motives not on the discovery but on the preservation of ideas — motives such as political objectives, superiority attitudes, and inertia. The treatment indicated here is the obvious one of encouraging (not only in others but in ourselves) the old virtues of modesty, tolerance, and intercommunication, and of diminishing the force of the extraneous motives or of channeling them to invention rather than apologia. An economist may be judged as reasonably healthy to the extent that his working problem, in his actual operations, includes not only the construction of a positive system on the basis of his working assumptions but also an analysis of the alternative working assumptions of rival schools.

The field of economic theory, during most of its independent existence, has been and still is the scene of a titanic battle of three powerful opponents (expressing different combinations of the theoretical, apologetic, reformist, and empirical attitudes) who are locked in mighty struggle with their own individual working problems but who, for all the energy expended, seldom come to grips with each other. The broad area of indecisiveness in this battle provides a constant challenge to economists to improve the efficiency of scientific method as applied to their discipline, to diminish their own insular tendencies insofar as scientific method permits, and to formulate a working problem which is broad enough to accommodate their rivals on a common footing with themselves yet concrete enough to bring into focus their significant differences and to point the direction in which these differences may be resolved.

GEORGE J. SCHULLER.

WASHINGTON, D. C.

# THE GENERAL THEORY OF LOCATION
## AND SPACE-ECONOMY[1]

### SUMMARY

I. Introduction, 476. — II. Early contributions to general location theory: the evolutionary approach of Alfred Weber, 479. — III. Predöhl and the substitution principle in location analysis, 482. — IV. The contribution of Hans Weigmann, 489. — V. The Lösch model of a space-economy operating under conditions of monopolistic competition, 494. — VI. The inter relation of trade and location theories, 502. — VII. Conclusion, 505.

## I

"The difficulties of the problem depend chiefly on variations in the area of space, and the period of time over which the market in question extends; the influence of time being more fundamental than that of space."[2]

Thus spoke Marshall, in line with Anglo-Saxon tradition, and in the half-century to follow, Anglo-Saxon economists were to hearken to his cry. Theoreticians of today are chiefly preoccupied with introducing the time element in full into their analyses, and the literature abounds with models of a dynamic nature. Yet who can deny the spatial aspect of economic development: that all economic processes exist in space, as well as over time? Realistically, both time and space must be vital considerations in any theory of economy. Unfortunately, however, aside from those of the monopolistic competition school of thought, particularly Chamberlin,[3] the architects of our finest theoretical structures have intensified the prejudice exhibited by Marshall. They continue to abstract from the element of space, and in doing so they are approaching a position of great imbalance.[4]

1. The author is indebted to the Social Science Research Council for financial assistance, and to Professors A. P. Usher, G. Haberler and E. H. Chamberlin for many helpful suggestions.

2. Alfred Marshall, *Principles of Economics* (8th ed.; London, 1936), Bk. V, chap. XV, sec. 1.

3. E. H. Chamberlin, *The Theory of Monopolistic Competition*, 1933; and Chamberlin's doctoral dissertation deposited under the same title in the Harvard University library, 1927. In his doctoral dissertation Chamberlin treats the space factor somewhat more thoroughly and more as an integral part of his theory. See also S. Enke, "Space and Value," this *Journal*, LVI (August 1942), pp. 627–37.

4. Outside of the field of monopolistic competition there have been scattered treatments by Anglo-Saxon theorists of certain aspects of space as an economic factor. For example, F. A. Fetter has treated space in "The Economic Law of Market Areas," this *Journal*, XXXVIII (May 1924), p. 525; treatises on international trade have attributed some importance to spatial resistances; rent theorists have been forced to recognize, however inadequately, the existence of

Let us consider, as an example, modern general equilibrium theory. This theory has not been molded in as dynamic a setting as have the several forms of partial analysis. Nevertheless, the latest contributors have concentrated their efforts on attacking the problem of time to the exclusion of that of space. Hicks,[5] Mosak,[6] Lange,[7] and Samuelson,[8] to name a few, have all treated an economy in which all factors and producers, commodities and consumers are, in effect, congregated at one point. Hicks, to be sure, begins by formulating the problem in a manner pregnant with spatial implications:

"It turns out, on investigation, that most of the problems of several variables, with which economic theory has to concern itself, are problems of interrelations of markets. Thus, the more complex problems of international trade involve the interrelations of the markets for imports and exports with the capital market. ..."

" ... The method of General Equilibrium, which these writers (Walras, Pareto and Wicksell) elaborated, was especially designed to exhibit the economic system as a whole, in the form of a complex pattern of interrelations of markets. Our work is bound to be in their tradition, and to be a continuation of theirs.[9]

But actually he confines himself to a wonderland of no dimensions. Apparently he assumes markets to be perfect, one price ruling throughout each of them. Or, otherwise expressed, transport costs and other costs involved in movement within a "market" are assumed to be zero. In this sense the factor of space is repudiated, everything within the economy is in effect compressed to a point, and all spatial resistance disappears.[1]

space in the separation of immobile natural resources and markets; and so forth. But in these latter instances only passing attention has been given to this vital consideration.

5. J. R. Hicks, *Value and Capital* (Oxford, 1939).

6. Jacob L. Mosak, *General Equilibrium Theory in International Trade* (Cowles Commission Monograph No. 7, Bloomington, Ind., 1944).

7. Oscar Lange, *Price Flexibility and Employment* (Cowles Commission Monograph No. 8, Bloomington, Ind., 1944).

8. Paul A. Samuelson, *Foundations of Economic Analysis* (Cambridge, Mass., 1947).

9. *Op. cit.*, p. 2. The words in parentheses are added.

1. The explanation may lie in Hicks's sweeping rejection of monopolistic competition theory generally in favor of perfect competition, on the ground that the former introduces elements of indeterminacy, whereas his preference is for determinate solutions (*op. cit.*, pp. 83–85). It is clear that Hicks fails to appreciate the spatial aspect of monopolistic competition theory; also the extent to which determinate solutions are obtained in monopolistic competition analysis through its consideration of the space factor.

Professor Schumpeter has pointed out to me in conversation that one might maintain that transport cost is implicitly contained in production cost, and that the Hicksian analysis is thus sufficiently comprehensive. My point is this: production theory, having gone beyond the mere statement that the producer

The approach to absurdity which is inherent in such a treatment is best illustrated by Mosak's work, *General Equilibrium Theory in International Trade* — which is excellent in other respects. Although Mosak expands Hicks's analysis to embrace an international economy, spatially speaking he is still dwelling within a dimensionless habitat. His study of the effects of international exchange, of unilateral payments, and of impediments to international trade,[2] can be interpreted as treating an anomalous and ridiculous field: *a one-point world, which somehow or other is conceived as divided into n parts, representing n nations, between which trade and trade barriers exist.*[3]

We may now consider the relations of general equilibrium theory to the general theory of location and space-economy envisaged in this paper. We conceive the general theory of location and space-economy to be one which comprehends the economy in its totality. Not only are the mutual relations and interdependence of all economic elements, both in the aggregate and atomistically, of fundamental importance; but the spatial as well as the temporal (dynamic) character of the interrelated economic processes must enter the picture. Seen in this perspective, Hicksian general equilibrium analysis is but a very special case of a general theory of location and space-economy, which concerns itself with the local distribution of factors and resources as well as with local variations in prices, and thus with the immobilities and spatial inelasticities of factors and goods.

In the sections to follow we shall restate and critically evaluate the contributions of the several authors who have pioneered in our field. We hope to push their analysis somewhat farther and to indicate directions along which it would seem desirable to develop the general theory of location and space-economy. It will not be surprising to find that, with the exception of Chamberlin,[4] these

maximizes his profits (in which statement all production costs are implicitly treated), from a methodological standpoint cannot justifiably treat certain production costs explicitly and other important ones implicitly in order to avoid the obstacles to analysis which the latter present. For a balanced treatment, the particular effects of transport and spatial costs in separating producers from each other must be considered. They are too vital to be sidestepped through implicit treatment, as has been done by Hicks and others.

2. Transport costs are not explicitly treated. The analysis of their effects, it is maintained, follows similar lines to that of the effects of import and export taxes (*op. cit.*, pp. 64–65).

3. At this stage in the development of theory, it is as unjustifiable and inadequate to lump transport costs into one category along with all trade resistances in the theory of trade, as it is to treat them as implicit production costs in the theory of production.

Although these indictments are levelled specifically at Mosak, they apply to others as well.

4. However, even Chamberlin does not attack the location factor *per se*,

authors have come under the influence of German thought. The classical school and their followers were too prone to overlook the local differences within the English economy. England's dominant international position and the dynamic aspects of her industrial development further helped to cloud their vision. It was in international trade theory that the spatial structure of the domestic economy was most explicitly assumed away or relegated to the background. This step facilitated a macroscopic process analysis (though quite elementary) of international trade which seemed so urgent to the classical school.

On the other hand, the reaction of German thought to classical teachings, which precipitated the rise of the German historical school, ploughed the ground for contributions in the field of "Raumwirtschaft." In the study of the stages of economic development the spatial structure of economic processes was necessarily a primary concern. And with the impress of the Lausanne school of thought upon German economics, it was almost inevitable that attempts would be made at a fusion of space with general equilibrium analysis.

## II

The first attempt to construct a general location theory is to be attributed to Alfred Weber in his Chapter VII, "Manufacturing Industry Within the Economic System."[5] It is true that the father of location theorists, von Thünen, who was far in advance of his time, did progress somewhat toward a general locational analysis. It may have been that his interests and experiences in the operation of his estate 'Gut Tellow' served to restrict the generality of his abstract thinking. Nonetheless, the seeds for developing the basic

but treats it as one of the leading manifestations of the broader category of "product differentiation." From his analysis emerges explicitly the need for applying the techniques of monopolistic competition in handling the space-economy of reality.

5. *Über den Standort der Industrien* (Tübingen, 1909); English translation with introduction and notes by Carl J. Friedrich, *Alfred Weber's Theory of the Location of Industries* (Chicago, 1929). The material in the rest of Weber's book does not concern general location. It deals with what is usually conceived of as of Weberian location theory, namely, an industrial location theory under the special conditions that: (1) the location and the size of the places of consumption are fixed; (2) the location of the material deposits is given; (3) the geographic cost pattern of labor is given, and at any one point the labor supply at constant cost is unlimited.

Weber's other important contribution ("Industrielle Standortslehre: Allgemeine und kapitalistische Theorie des Standortes," *Grundriss der Sozialökonomik*, Part VI (2d rev. ed.; Tübingen, 1923) merely touches the field of general location theory.

methodology in analysis of specific as well as general location problems can be found in Thünen's work.[6]  The science of economics has suffered from the relative neglect of his methods during the nineteenth and early twentieth centuries.

Launhardt, the other major predecessor of Weber, also failed to achieve sufficient generality in his analysis.  In fact Launhardt's studies of industrial location and market areas[7] treated a much narrower set of circumstances than were encompassed in Thünen's isolated state.

Weber's attempt at general locational analysis was undoubtedly greatly influenced by the writings of Roscher and Schäffle.[8]  Weber pursued an essentially evolutionary approach.  He tried to develop the general basis upon which any given historical system orients itself, or, in other words, a theory of the transformation of locational structures.

His method is to inquire into the forces that come into operation, when a people occupy an undeveloped country and establish an isolated economic system.  At first an agricultural stratum forms to produce the necessary means of subsistence.  The settled area with its agricultural population serves then as the geographical foundation for all other strata.  It determines in the first instance the locus (places) of consumption for the second stratum, namely, the primary industrial stratum, which produces for the agricultural stratum.  In turn, the primary industrial stratum serves as the geographical sphere of consumption for the third stratum, namely, the secondary industrial stratum.  This third actually consists of numerous substrata, each of which is oriented to and is smaller than the preceding one, the first being the only one directly oriented to the primary industrial

6. Johann Heinrich von Thünen, *Der isolierte Staat in Beziehung auf Landwirtschaft und Nationalökonomie* (Hamburg, 1826).  See also the interesting article by Bertil Ohlin, "Some Aspects of the Theory of Rent: von Thünen vs. Ricardo," *Economics, Sociology and the Modern World: Essays in Honor of T. N. Carver* (Cambridge, Mass., 1935).

7. See, in particular, "Die Bestimmung des zweckmässigsten Standortes einer gewerblichen Anlage," *Zeitschrift des Vereins deutscher Ingenieure* (Berlin, 1882), Vol. XXVI, No. 3, and *Mathematische Begründung der Volkswirtschaftslehre* (Leipzig, 1885), Part III.

8. Wilhelm Roscher, "Studien über die Naturgesetze, welche den zweckmässigen Standort der Industriezweige bestimmen," *Ansichten der Volkswirtschaft aus dem geschichtlichen Standpunkte* (3d ed.; 1878); A. Schäffle, *Das gesellschaftliche System der menschlichen Wirtschaft* (3d ed.; Tübingen, 1873).  Both were of the German historical school and were primarily concerned with discovering whether or not there were any natural laws or regularities in the evolving locational structures of economies.  Their contribution rests in their collection of historical facts and in their presentation of an abundance of conflicting ideas.

stratum. These three strata form the core of the economic system. The mass of local tradesmen and functionaries, engaged in the process of circulation and in performing personal services, strengthens proportionally the different parts of this system.

A fourth, the central organizing stratum, is essentially independent of any of the three preceding ones. It consists of officials and businessmen with general organizing and managing functions, of members of the liberal professions, and of persons living off accumulated wealth. Their pattern of locations within the economic system, if not arbitrary, is determined not by economic forces but by others. A fifth, the central dependent stratum, is formed and tied to the central organizing stratum, in the same way as is the secondary industrial to the primary industrial stratum.

The locational structures of these five strata are interrelated with forces playing back and forth among them. For example, though the agricultural stratum appears on the scene first, the formation of cities incident to industrial development induces rearrangements of the agricultural structure to conform more closely to the pattern of concentric zones as conceived by von Thünen.

This is as far as Weber goes. Despite the later writings of Engländer[9] and Ritschl,[1] which elaborate and develop in other directions the evolutionary approach, no essential advance in this technique of general analysis is made. The technique is inadequate; it does not present any general, heuristic principle by means of which one can order the spatial complexities involved in the total location of economic activities. It merely records the interrelations of the various strata, and some of the reactions of one stratum upon another.

9. Oskar Engländer, "Kritisches und Positives zu einer allgemeinen reinen Lehre vom Standort," *Zeitschrift für Volkswirtschaft und Sozialpolitik*, Neue Folge, Vol. V, Nos. 7–9, 1926. With Engländer the problem is to investigate first the spatial form of primary production, i.e. of a land and forest economy, where all households are self-sufficient. Next, specialized products are assumed to be cultivated on land of particular quality, and the consequent changes in the spatial structure of the economy are observed. In turn agricultural industries, mining, manufacturing and other economic elements and complicating factors are successively introduced and the resulting spatial realignments of economic activities noted. Ultimately, according to Engländer, an approximation to the picture of a modern economy is realized.

1. Hans Ritschl, "Reine und historische Dynamik des Standortes der Erzeugungszweige," *Schmollers Jahrbuch*, LI (1927), 813–70. Ritschl, recognizing that the location picture is historically relative, follows Weber's classification of strata and traces in detail their development during the periods of village, city, territorial, national, and world economy.

Also see R. G. Hawtrey, *The Economic Problem* (New York, 1925), chaps. vii and ix, in which to some extent he adopts the approach of Engländer and Ritschl.

For any given stratum, or combination of strata, it fails to get at the rule or rules governing structure and provides no common denominator in terms of which all the forces stemming from the various interrelations can be expressed and evaluated, and by means of which a net effect could perhaps be deduced. This is the task of a general theory of location and space-economy.[2]

Nonetheless, the evolutionary approach is very useful. It not only furnishes a convenient and meaningful breakdown for studying historical sequences of locational structures and for classifying historical facts, but also will be very suggestive for pursuing dynamic analysis, once an adequate general static theory has been achieved.[3]

## III

Shortly after the appearance of Weber's book, Bortkiewicz[4] and Schumpeter[5] recognized the need of a general equilibrium analysis to supplement partial locational theories.[6] Considerably later Engländer[7]

2. This is clearly indicated in Ritschl's work. After describing in detail, in a section on pure dynamics, the various possible effects of changes in ten or more major locational elements, he is unable to find a method for combining their effects. (*Op. cit.*, pp. 853–56.)

3. In the order of treatment of subject matter the work of E. M. Hoover, *Location Theory and the Shoe and Leather Industries* (Cambridge, Mass., 1937) somewhat resembles the evolutionary approach. But Hoover's analysis is definitely partial, though in a broad setting. Through drawing up carefully a set of assumptions and relaxing them one by one, he is able to proceed from an analysis of extractive industries to a treatment of manufacturing first under simple conditions and then under more complex ones. He emphasizes the major specific forces at work and does not pay too much attention to general interrelations, especially when they can be stated only in broad terms. In this way he is able to synthesize the various theoretical contributions of his predecessors that are of practical value and by employing illustrative empirical material is able to stick close to reality. From the standpoint of balance and sound judgment Hoover's writings are the best. See also his *Location of Economic Activity* (New York, 1948).

The other major work in English on location theory again follows a partial approach both in its static analysis and in its dynamics where it relies upon extensive use of historical material (W. H. Dean, Jr., *The Theory of the Geographic Location of Economic Activities*, doctoral dissertation, Harvard University, 1938. Selections published by Edward Brothers, Inc., Ann Arbor, Michigan, 1938).

4. *Deutsche Literaturzeitung*, XXXI (1910), 1717–24.

5. *Jahrbuch für Gesetzgebung, Verwaltung und Volkswirtschaft*, XXXIV, No. 3 (1910), 444–47.

6. V. Furlan ("Die Standortsprobleme in der Volks- und Weltwirtschaftslehre," *Weltwirtschaftliches Archiv*, II (1913), 1–34) makes a somewhat abortive attempt at general locational analysis. The complicated interrelations of various economic factors as well as the "spatial transformation of goods" are fully recognized; but the contributions to knowledge are essentially along the lines of developing overly-simplified models of markets, domestic and international — more specifically of determining points of collection and distribution of goods, of export and import and the related paths of commerce.

7. *Op. cit.*

came to appreciate in full the implications of a general theory of location. The pure theory of location, according to Engländer, is the general theory of "local conditionality" within an economy. Any given entrepreneur, in choosing the site at which to produce or render services, considers the various supply prices existing in the various localities for the inputs that he might possibly employ; at the same time he considers the various prices which might be obtained in the various localities for his product or services. When finally he does locate at a site, he influences in turn the prices of various inputs and outputs. Through being so interrelated, the pattern of local price differences and the location of economic activities are simultaneously determined by a general theory of "local conditionality."[8]

Somewhat earlier than Engländer's publications there appeared an article, "Das Standortsproblem in der Wirtschaftstheorie,"[9] by Andreas Predöhl (1925), which utilized a principle by means of which a general equilibrium approach could be systematically applied to location analysis. This was none other than the familiar substitution principle, already well established in general equilibrium theory.

Although Predöhl did visualize new horizons in the extended use of this principle, he tried unfortunately to remain within the scope of traditional thought. He purported to deduce a general location theory as a special case of the existing general economic theory, as a logical and inherent element of it. The general economic theory to which he alluded was the theory of interdependent prices, of general equilibrium as expounded successively by Walras, Pareto and Cassel. He wished to investigate how far the location problem is a price problem; location theory, a price theory. In other words, to what extent does the local distribution of production lie inside the economic relationship of interdependent prices?[1]

8. Further (1) by classifying raw materials and factors of production, whether mobile or immobile, as place-free (available everywhere under the same conditions), conditionally place-bound (available at all or some places, under unequal conditions) and unconditionally place-bound (present at one site), and (2) by conceiving immobile goods as goods of infinite weight which enter into production with infinite weight-loss, Engländer brings together the specific location theories of industry and agriculture within the confines of his pure location theory, not as distinct compartments, but as internally related sectors.

Elsewhere, too, Engländer has attacked a broad range of location problems, but only through elucidation of simplified, isolated cases. See his *Theorie des Güterverkehrs und der Frachtsätze* (Jena, 1924), and "Standort" in *Handwörterbuch der Staatswissenschaften* (4th rev. ed.; Jena, 1926), VII.

9. *Weltwirtschaftliches Archiv*, XXI (1925), 294–331. A briefer article in English, "The Theory of Location in its Relation to General Economics," appeared in the *Journal of Political Economy*, XXXVI (1928), 371–90. Also see Predöhl's reply to Engländer's criticism, "Zur Frage einer allgemein Standortstheorie," *Zeitschrift für Volkswirtschaft und Sozialpolitik*, V, Nos. 10–12 (1927), 756–63.

1. "Das Standortsprobleme . . . " *Op. cit.*, pp. 295–97.

Predöhl contends that the problem of the local distribution of economic activity is synonymous with the problem of the distribution of determined groups of production factors (he groups production factors under the categories of land, labor and capital), since every economic activity uses a grouping of factors. The distribution of determined groups of production factors in turn is a special case of the distribution of productive factors in general.[2] To Predöhl, general interdependence theory explains the distribution of productive factors in general by means of the principle of substitution. Therefore general location theory is deducible from the application of the principle of substitution to the employment of the several groups of productive factors.[3]

Predöhl in his reasoning overestimates the scope of Walrasian-Casselian general equilibrium analysis. On the whole he seems to be under the impression that it implicitly embraces the space element in its entirety. However, as indicated previously, modern as well as earlier general equilibrium analyses, with minor exceptions, concern a one-point world. The element of transport cost is generally abstracted; factors and products possess perfect mobility. In essence there is no spatial distribution of factors; the relevant problem is the distribution of factors among the various types of production. In reality, then, the situation is the reverse of what Predöhl has conceived, although several times he appears to realize the truth of the matter. As we have pointed out, Walrasian-Casselian general equilibrium analysis is but a special case of a general location theory.

Nevertheless, the tools shaped by general equilibrium theory are useful, as Predöhl discovered. Starting with the familiar case of Thünen's isolated state,[4] Predöhl assumes all locations fixed except that of one enterprise. A shift of this enterprise toward the periphery implies that capital and labor outlays (including transport outlays) are substituted for land use outlays. The reverse takes place in a shift toward the central consumption point. Application of the principle of substitution will yield the site of minimum cost so far as these two all inclusive groups of expenditures are concerned. How-

2. As will be shown later, this statement is very weak, if not untenable.
3. *Ibid.*, pp. 299–303.
4. The features of Thünen's familiar model are: a uniform plain with equal fertility and possibilities for agricultural production at all points, at the center of which lies a city possessing potential transport facilities of similar character in all directions (i.e. transport costs proportional to weight and distance). Production aligns itself around the city in rings in accordance with the price and transport cost of each particular product cultivated. Predöhl adopts at the start an expanded version of Thünen where all conditions for all production, whether agricultural or industrial, are uniform throughout the plain (*ibid.*, p. 299).

ever, within these two all inclusive groups there are other substitution points. For example, within the former group, there is a substitution point between transport outlays and *local* capital and labor outlays (such as is involved in determining whether or not to process a product in order to reduce its weight or bulk); and within the category of transport outlay, there may be a substitution point involved in allocating a given portion between transporting a raw material lying at the periphery and transporting a raw material lying near the consumption center. In this manner innumerable interdependent points of substitution arise which determine the location of any individual enterprise. This proposition, states Predöhl, can be extended by means of general equilibrium analysis to cover the location of all economic activities.[5]

Inequalities in local resource patterns, land, labor, capital and transport do not invalidate the operation of the substitution principle. They present various technical possibilities for production which are different from those that would exist in Thünen's homogeneous plain; but essentially these new production possibilities, like the old, can be expressed in terms of economic values, and thus fall within the scope of substitution operations. Similarly, economic values can be imputed to various historical-political forces, though here many more difficulties and arbitrary elements creep in. Recognizing these various limitations (for example, in accounting for the locus of consumption of the rentier classes) Predöhl does, however, maintain that the locationally relevant substitution points, thus logically deduced, are applicable in general.[6]

It is to be expected that Predöhl in this first attempt at substitution analysis would be unable to resolve all the difficulties that beset his path. His argument is particularly weak when he becomes specific and illustrates substitution operations — a step which he avoids as much as possible. For example, he utilizes a vague concept, namely, a land use unit, and speaks of rent outlays at different sites as being proportional to the quantities of land use units at those sites. Land more distant from a city and yielding less rent therefore relates

5. "Der Standort der Produktion bzw. Produktionsstufe ist also bestimmt durch ein System von Substitutionspunkten, das derart gegliedert ist, dass die Gruppen einer übergeordneten Kombination untergeordnete Kombination in sich enthalten. Übertragen wir diese Lösung auf sämtliche Produktionen, dann können wir unter Zuspitzung eines allgemeinen Casselschen Satzes auf unser besonderes Problem sagen: Wenn das Preisverhältnis in dieser Weise für jeden einzelnen Betrieb die standortlich relevanten Substitutionspunkte bestimmt, sind offenbar durch dasselbe für die gesamte Gesellschaft die zu verwendenden Mengen im Verhältnis zueinander, mithin die Standorte bestimmt." *Ibid.*, pp. 306–7.

6. *Ibid.*, pp. 308–11.

to fewer technical units of land use than land less distant which yields greater rent. Engländer easily demonstrated that this proposition is false: that two pieces of land unequally distant from a city can be of the exact same quality and be utilized to the same degree, and yet yield different rents.[7]

Further, Predöhl tends to convert all spatial and quality differences into differences in quantities of use units. Immobile labor, situated at diverse places and of different qualities, can be converted into amounts of labor use units and thus made comparable. And so with all types of resources.[8] In this way all geographic differences in land, labor and capital can be summed up into use units of land, labor and capital at any given point. This reasoning lies behind Predöhl's argument that the distribution among various economic activities of determined groups of production factors (each group at any point of time having a unique spatial position) is a special case of the distribution among various economic activities of productive factors in general in terms of a one-point society.

It is not necessary to carry the argument to such an extreme, if not untenable, position. No further development of Predöhl's stand has been published. It is appropriate now to revise and extend his model in order to strengthen the basis for an extensive use of the substitution principle in location analysis.

First one ought to distinguish between two types of substitutions: that between distance inputs and that between outlays. If there is any sense at all to location economics, it is because there are certain regularities in the variations of costs and prices over space. These regularities arise primarily because transport cost is some function of distance. If this were not so, if transport costs were completely irregular and their changes unpredictable — if for example transport costs on a certain item were positive for a distance of 100 miles, and negative for a distance of 101 miles — there would be little sense in discussing the economics of plant locations. The spatial pattern of industrial concentrations, of consuming centers, and of the production of raw materials would be quite arbitrary from the economic standpoint.[9]

Since it is the distance factor that is the heart of locational analysis, there is every reason to speak of distance inputs, and of transport rates as prices of those inputs. Location theorists unfor-

7. "Kritisches und Positives . . .", *op. cit.*, pp. 499–500. See Predöhl's weak reply ("Zur Frage . . . ", *op. cit.*, pp. 758–60).

8. "The Theory of Location . . . ," *op. cit.*, pp. 380–81.

9. The points in this and the following paragraphs of this section are developed at length in a forthcoming study.

tunately have shied away from such a concept. However, it brings into bold relief the basic aspects of spatial analysis without the necessity of tagging each unit of land, labor and capital with a set of spatial coordinates or of converting them into common units, if, indeed, this can be done. The problem of production becomes a problem of choosing the right combination of the various types of capital, labor, land and distance inputs. In terms of the Weberian locational triangle whose three corners represent a source $A$ of one raw material, a source $B$ of another raw material and the consumption center $C$, we have three distance inputs to consider (whose prices are the transportation rates on the raw materials and finished product): distance from $A$, distance from $B$ and distance from $C$. The actual combination of distance inputs is of course interrelated and simultaneously determined with the amounts of the various types of land, labor and capital inputs.

The selection of the correct substitution points between distance inputs is easy to visualize, although in practice it may be difficult to effect because of the complicated nature of transport rate structures. However, selection of the correct substitution point between a distance input and a labor input, or between the two groups, distance inputs and labor inputs, cannot be so satisfactorily handled. There is too little spatial regularity in the variation of labor costs, although to be sure one can speak roughly of regional wage levels and of the tendency for money wages to rise or fall with distance from an urban center, depending upon the type of region under consideration. Extra-economic, cultural factors too often are dominant in determining the spatial pattern of wages. Its very irregularity makes impossible any such type of substitution analysis as was discussed with respect to distance inputs.[1] The choice of the optimum location thus requires an outright comparison of outlays on the various kinds of labor, or of total labor outlays and total transport outlays, or of total labor outlays and total interest outlays, and so forth. Substitution analysis in terms of outlays[2] must supplement substitution analysis in terms of distance inputs in order to achieve a proper locational methodology.

It is interesting to examine how Weber's industrial location theory[3] falls into the scheme of substitution analysis. We find that

1. To the extent that there is regularity in wages over space which is related to differences in transport cost in obtaining the goods in the laborer's market basket, then substitution analysis in terms of distance inputs is still relevant. Compare Hoover, *Location Theory and the Shoe and Leather Industries*, chap. IV.

2. And in terms of various types of revenues if more than one commodity is produced by a firm.

3. That is, the theory contained in the first six chapters of his book (*op. cit.*).

this theory can be broken down into a general, common core of substitution points. As already intimated, Weber's optimal transport point (point of overall minimum transport costs) represents nothing but the correct combination of distance inputs. The possibilities of deviation to labor locations (labor orientation) are indicative of discontinuous substitution points between transport outlays and *local* labor outlays. Weber's agglomeration tendencies are potentialities for substitution between various types of outlays, chiefly between outlays on land and outlays on labor, capital or transport or any combination of these three.[4]

Thus by the above approach, Predöhl's original conception of substitution analysis can be made more digestible and broadened into a general equilibrium theory of space-economy, which includes the actual modern general equilibrium theory as a special case.[5]

4. Compare Predöhl, *op. cit.*, pp. 311–20.

Predöhl and others have rightly pointed out that Weber's industrial location theory is chiefly based on technical empirical knowledge. Transport costs are reduced to weight and distance, i.e. to technical factors; varying raw material prices and other elements are reduced in similar fashion. Technical concepts such as locational weight, material index, coefficient of labor, Formkoeffizient and others are the critical measures. The point of transport orientation is merely the point of minimum transportation in terms of *ton-kilometers*. Essentially Weber abstracts from most economic interrelations and reactions. Only under severe limitations is Weberian doctrine generally applicable.

However these criticisms do not detract from the merits of Weber's contribution. Formal theory, in and of itself, is highly unsatisfactory, too general and, accordingly, too sterile. As Predöhl emphasizes, it needs to be supplemented by concrete information; abstract and vague values must be replaced by exact, quantitative data. In other words, supplementary explanations are required, even if they are obtained in such a manner as to limit their general validity. It is in fulfilling this need that Weber's work is of great significance. If the general theory of location constructed upon the principle of substitution is to be of pragmatic value, to it must be added empirical location theory and statistical investigation which seeks out regular movements in major economic variables, even though this means eliminating the numerous special factors which affect each individual situation. Weber's theory of industrial location is just such a supplementary, empirical theory (it excludes economic details which he considers relatively unimportant, and thus in great part hypothesizes that the set of technical substitution points approximately parallels the set of economic substitution points). But ultimately all such empirical technical functional observations must be translated into economic terms.

Indeed, it is only by utilizing chiefly the Weberian approach with supplementary economic data that the writer has found it meaningful to analyze the locational structure of the iron and steel industry. See his "Some Locational Factors in the Iron and Steel Industry since the Early Nineteenth Century," *Journal of Political Economy*, LVI, June 1948; and (with W. Capron) "The Future Locational Pattern of Iron and Steel Production in the United States," *Journal of Political Economy*, LVII, April 1949. Also see E. Niederhauser, "Die Standortstheorie Alfred Webers," *Staatswissenschaftliche Studien*, XIV (Weinfelden, 1944).

5. The general pervasiveness of the fundamental substitution operation can

## IV

In this section we digress from the main argument to present a summary of the contributions of Hans Weigmann,[6] which have received but slight attention in the literature. Weigmann's writings on general location theory are very difficult to comprehend, both because of his vague style and the complexity of the basic concepts. These concepts do not lend themselves to a general synthesis as do those of other contributors. Nonetheless, they seem to disclose some of the most promising channels of exploration for further theoretical development.

Weigmann attempts to formulate the foundations for a *realistic* economic theory which embraces the spatial structure of economic processes, the spatial extent and bonds of markets, the spatial inter-relations of all economic quantities.

The first principle that Weigmann establishes is that a theory of space-economy embraces a theory of limited competition. Actually all factors and goods, regardless of setting, face immobilities in all directions, of varying extent; and in accordance with the nature of the obstacles to movement, whether they be economic, social, political or cultural, markets are restricted in scope. The competition which

be demonstrated in another way. Commodities (including all types of factor inputs and products) may be classified as follows: mobile or immobile; obtainable at a single source or obtainable at many; dispensable or indispensable; pure or weight-losing (i.e. inputs which add all or just part of their weight to the final product). Only in the case of an immobile, single source, indispensable commodity, whether pure or weight-losing, is it possible that there be no substitution operation in the location process. It should be noted, however, that few commodities, if any, are indispensable from the social-aggregative standpoint. A dispensable commodity implies a substitution point between at least two commodities and, from a locational standpoint, between the distance inputs and outlays associated with the location of the given commodity (whether it be a deposit of a raw material or a site of production) and those associated with the location of a potential replacement. A commodity available at many sources implies substitution in the use of the several sources. Mobility introduces substitution with respect to the various sites to which a good can be transported. And weight loss brings out substitution relations between various possible distance inputs whose costs constitute total transport outlay.

That category which includes commodities which are mobile, present at many sources, dispensable and weight-losing involves the broadest range of substitution effects. Any other category of commodities can be viewed as a special case of this one, in which the substitution potentialities are more limited.

6. "Ideen zu einer Theorie der Raumwirtschaft," *Weltwirtschaftliches Archiv*, XXXIV (1931), 1–40; and "Standortstheorie und Raumwirtschaft" in *Joh. Heinr. von Thünen zum 150 Geburtstag*, ed. by W. Seedorf and H. Jurgen (Rostock: Carl Hinstorffs, 1933), pp. 137–57. To trace the development of Weigmann's thought the reader is also referred to the following of his works: *Kritischer Beitrag zur Theorie des internationalen Handels* (G. Fischer: Jena, 1926), and *Politische Raumsordnung* (Hanseatische Verlagsanstalt: Hamburg, 1935).

any good or factor can offer to other goods and factors at different locations is incomplete. The existence of physical space implies immobility, limited competition and spatial inelasticity (or negative spatial elasticity). Thus the generally accepted principle of pure competition is not applicable to the analysis of spatial economic processes.[7]

A second basic principle concerns the question of form. In place of the customary linear, causal analysis Weigmann favors the approach of general equilibrium theory in the employment of Gestalt analysis. He observes the space-economy as a whole in its full array of spatial markets. In that sense he aims at presenting a realistic, functional picture of the "form-full" of economic life, wherein the various elements are weighted in accord with their importance. Having adopted this methodology, he confronts the primary problem of determining the basic form (Grundgestalt) of economic phenomena, i.e., the Gestalt core. This basic form should then provide an heuristic principle to help master and order systematically the "form wealth" of real economic life, or in other words, the countless spatial forms of moving economic processes.[8]

At this point Weigmann differentiates between statics (mobility or competition as potential energy) and dynamics. Since he purports to describe the space-economy in its realistic setting, he is compelled to complicate his problem manifoldly by introducing the time element, by assigning time coordinates to his various markets and processes. Weigmann poses the perplexing problem of dynamics as follows: to choose that time period which would yield in the resulting spatial array of markets a competition field (a broad market area in time and space) which could be valid as the basic form.[9] He resolves the problem by formulating a concept quite difficult to comprehend, the concept of "relative maximum." It states that as an increasing amount of physical space (therefore spatial resistance) is to be overcome in movement by an economic object, the time period necessary for such movement increases until it reaches a maximum — a max-

---

7. "Ideen zu . . . ", *op. cit.*, pp. 6–9. These points are particularly well developed by Chamberlin in his doctoral dissertation of 1927, *op. cit.*, especially pp. 105–09, 167–84; and in *The Theory of Monopolistic Competition*, 1933, *passim*. In the latter, a portion of the earlier analysis devoted to two-dimensional space was simplified to one dimension and removed from chapter 5 to appendix C.

8. *Ibid.*, pp. 9–12.

9. " . . . welches Konkurrenzgebiet ist essentiell im Sinne des Gestaltganzen, wenn bei beliebiger Absteckung der Zeitgrenzen eine Fülle räumliche variierender Flächen entsteht? Oder anders ausgedrückt: welcher Zeitraum ist zu wählen, damit ein mit diesem Zeitmass gegebenes Konkurrenzfeld als Grundgestalt gelten kann?", *ibid.*, p. 14.

imum in the sense that given still more time a further spatial movement would be improbable because of the overpowering force of the countless obstacles. There, where the time period reaches its maximum, competition ends and the competition field becomes bounded. In other words, the force of competition does not have the power to span a distance greater than the radius (or axis) of its field, irrespective of the time factor for all practical purposes. This principle contains the definition of basic form. The basic form is depicted as that unit of space (corresponding to a market region or competition field) of the relatively greatest time-weight — hence of the relatively greatest stability and permanence.[1]

Having exposed the tremendous magnitude of the task of formulating a theory of space-economy, Weigmann stops for breath. How to locate the basic form? How to represent as an empirical Gestalt unit the multitude of interlaced, mutually related individual markets, market strata and market densities? From here on our author can only offer fruitful suggestions and preliminary observations for conquering the manifold difficulties which appear. First we have the classification of markets according to structure. Each individual commodity market including its labor, capital and land orientation possesses a particular structure which offers a certain resistance to change. Some change frequently; others slowly. Some are active; others highly inactive. By definition those markets of a relatively permanent nature, of persistent inactivity, are grouped together as the essential ones, as the basic form; their combined structure determines the basic structure of the Gestalt whole, of the space economy under question. On the other hand, the rapidly changing markets are considered as accidental or secondary; their movements are characterized as minor modifications of the Gestalt form, and these movements are to a certain extent conditioned by the already determined basic form, by the core of markets of greatest continuity. Fundamental, organic change of the Gestalt picture of the space-

1. " ... und führen angesichts des Vorhandenseins eines relativen Maximums zeitkostender Bewegung innerhalb jedes Gestaltganzen zu einem Bilde sich bündelnder und überschneidender Konkurrenzfelder als den akzidentiellen und peripheren Erscheinungsformen einer zentralen Grundgestalt. Der Begriff des relativen Maximums besagt dabei folgendes: Die Konkurrenz wird gradweise beschränkt und als dort aufhörend bedacht, wo die zahlreichen hemmenden Faktoren den sukzessiv steigenden Zeitaufwand der Bewegung zu einem Maximum hinführen, bei dem unter den konkret gegebenen Umständen die weitere Bewegung unwahrscheinlich wird oder auf lange Sicht nicht mehr die Kraft besitzt, mit Ansicht auf anhaltenden Erfolg Spannungsunterschiede aufzugleichen. Die Grundgestalt ist also die Raumeinheit des relativ grössten Zeitgewichtes; die Dauer des Bestandes gibt ihr den Charakter der essentiellen Form." *Ibid.*, pp. 14–15.

economy, therefore, implies only change within this relatively immutable core of persistent markets.[2]

Our task is further illuminated by reference to the structures of the specific markets for land, labor and capital goods. Weigmann maintains that the markets for the productive factors of labor and land are primary constituencies of the basic form. Movements in all commodity markets course back to these two, whether directly or indirectly, through semi-finished products and various stages of production. And in these markets for labor and land which offer great resistance to change are focused the facts of scarcity within the economy.

The land market is portrayed as a spatially-connected area of supplied land services. Actually each individual piece of land is distinct and immobilized by nature, so that the market for it has no spatial extent. But for practical purposes Weigmann conceives of a Gestalt whole (space-economy) already in existence. This whole exerts an hypothetical aggregate demand, which in turn defines the boundaries of the land market, the peripheral area being considered as marginal land. The supply of land in general is not perfectly inelastic, spatially speaking. Change in the land market ensues (1) from additions or subtractions at the fringe to the land under cultivation, i.e., an expansion or contraction of the space base of the economy, and (2) from variations in the intensity of use and in the methods of cultivation and organization of each individual land unit.[3]

The size and nature of the hypothetical demand mentioned above is obviously related to the labor market. The labor market, in contrast to the market for land, is much less rigid and invariable; and for that reason its description and conception is theoretically and empirically much more difficult. There are many forms of labor immobility and inelasticity. Weigmann makes a beginning at analysis by explaining one, namely migration mobility. To delimit the labor market accordingly, one must recognize the various time stages of migration (e.g., seasonal, cyclical, and secular) and their spatial forms. In line with familiar Weberian technique, the long run labor base is presented as a continually moving, organic process, whereby labor step by step through varying intervals of time, gradually moves from farms or rural communities to giant metropolitan centers via town and urban clusters of increasing size.[4] This ever structural

2. *Ibid.*, pp. 16–19.
3. *Ibid.*, pp. 20–23.
4. Weber ("Industrielle Standortslehre ...", *op. cit.*, pp. 74–84) distinguishes two stages of modern capitalistic development: (1) Bound (gebundene) capitalism and (2) Free capitalism. In the former, which characterizes the six-

movement within the labor base is designated as one of the essential dynamic aspects of modern space-economics.[5]

With respect to markets for capital goods Weigmann offers a few suggestions. First we must distinguish capital in substance from capital in title. The former obviously has far more limited mobility. Second, capital goods (which in its broadest formulation includes all commodities) must be classified according to the extent to which they become bound up in production. At one end of the scale would be "combination-free" capital goods; at the other end would be capital goods permanently tied to a given production combination. The spatial elasticities and markets of the several divisions of capital goods would vary accordingly. Unfortunately, present day terminology regarding capital (e.g., fixed and circulating) is unsuited for depicting its spatial elasticity; nor has theory recognized the influence of spatial elasticity upon the various other elasticity forms of capital.

The formulation of the problem of a theory of space-economy thus is more comprehensive than that of traditional location theory. The latter has chiefly treated capital as a "combination-free" factor in its long run agglomerative setting, but has given little attention to the mobility of existing equipment, to the short run adaptability of capital goods. In fact, for an empirical theory, there are even strong grounds for considering the mobility of a given combination of various productive factors as a whole, rather than of their constituent parts — a phase of the problem which location theory has rarely posed.[6]

Thus Weigmann sketches his picture of the space-economy — as a rhythmic-moving Gestalt whole with a core composed of the markets for land, labor and capital goods and of numerous other markets superimposed upon these, overlapping and irregularly intersecting each other and at times extending into other space economies.[7]

teenth to the eighteenth centuries, labor is historically fixed, locally immobile. In the latter, which characterizes the present times, labor becomes mobile, released from its historical bonds. The economies of concentration and large scale organization can come into operation and can offer incentives for huge masses of labor to agglomerate at given points. On the other hand, these forces are offset by the community attachments (home feeling) of the individual laborer, by his lack of perspective and initiative and by the consequent increase in rent at the points of agglomeration. The net result is the step by step migration already mentioned.

5. "Ideen zu . . . ," *op. cit.*, pp. 23–27.
6. *Ibid.*, pp. 27–32.
7. In the Thünen Festschrift (*op. cit.*) Weigmann commences with existing location theory (the Engländer and Predöhl versions) and approaches a theory of space-economy ("total localization") in part through the extension of Predöhl's substitution principle to include "quantity elasticity." Quantity elasticity is synonymous with a broad definition of elasticity of supply, one that embraces,

His presentation lacks clarity and frequently one is forced to construe an imaginary model in order to follow the argument. Nevertheless, one obtains penetrating insight into the subtle spatial relations of economic life, and is given an original as well as a challenging view of the immense magnitude of the assignment.

## V

In Section III we developed the framework of a general (static) theory of location and space-economy in terms of the substitution principle. However, it has been customary in general equilibrium analysis to present the relations of a given one-point economic order by means of a system of mathematical equations. Should a solution for this system of equations exist, the merit of the presentation is generally regarded as considerably enhanced. Can a solvable system of equations be evolved also for a space-economy?

Tord Palander, in the first major work on location theory to originate outside of Germany, addressed himself to this question.[8] He considered insuperable the difficulties encountered by the general approach in representing or even closely approximating reality.

First, states Palander writing in 1935, the Walras-Pareto-Cassel general equilibrium theory in its present form is meaningful for a locational analysis only of an economic district wherein transport costs are zero, capital and labor perfectly mobile, and technical conditions of production uniform throughout — in other words, where the district in question can be compressed into a point market. To be sure, he continues, a somewhat closer approximation to reality can be obtained by withdrawing one by one the simplifying assumptions given above. For example, there might be introduced into the simplified model the following series of complications: freight costs on product based on distance and weight, transport costs for mobile production factors, equal real wages throughout the district, consumption as dependent upon location choice, and so forth. Even so, contends Palander, this procedure would not take us far, for in respects other than the neglect of local differences in demand and supply of factors and commodities, the deviations of a general equilibrium theory from reality are severe. Interdependence theory has as an underlying premise the principle of pure competition. Yet, in no sense at all, can the traditional interpretation of this premise hold

among others, spatial elasticity. In this way Weigmann brings out the logical bond of location theory and general price theory.

8. *Beiträge zur Standortstheorie* (Almqvist & Wiksells Boktryckeri-A.-B.: Uppsala, 1935), chaps. X and XI.

when we introduce space and thus transport costs into the analysis. If the various places in a region under consideration are treated as different markets (corresponding in this way to the varying local prices resulting from transport costs between these places) then the necessary condition of a large number of buyers and sellers for each commodity and factor at each market cannot be fulfilled. If the region itself is viewed as one market, one could interpret the different prices ruling for a given commodity at the various places within the region (1) as signifying nonuniformity of product, or better yet, (2) as signifying a uniform product in a persistently imperfect market where individuals are in monopoly situations in accordance with the advantages of their respective positions. Neither case could be regarded as pure competition.[9]

If one now discards the premise of pure competition, he must necessarily forsake certain supplementary simple principles which have served as scaffolding for general equilibrium theory, namely, that the price of a commodity equals cost (the latter including a normal profit) and that the price of a factor equals the value of its marginal product.[1]

Further censure of general equilibrium analysis follows from its limitation to static conditions — a widely recognized limitation which does not need to be discussed here.[2] Palander insists on the necessity of depicting the economic development process. His conscience thus compels him to forego Walrasian economics in favor of the Launhardt-Weber tradition. His energies are confined to analyzing the economic starting point, the adaptations of enterprise during a time period, the movement of factors during the same period, and the concomitant changes of technique, institutions and consumer base.[3]

August Lösch, however, has not accepted these views. In his monumental work, *Die räumliche Ordnung der Wirtschaft,*[4] he goes beyond partial analysis and the mere recognition of the complex spatial interrelations of economic factors. He presents succinctly

9. *Ibid.*, pp. 273–77. Here, again, excellent statements of these points are to be found in Chamberlin, *op. cit.*

1. *Ibid.*, pp. 277–78.

2. However, see the work of Samuelson which relates the comparative statical behavior of a general equilibrium system to its dynamical stability properties (*op. cit.*, Part II).

3. *Ibid.*, pp. 278–85.

4. G. Fisher: Jena, 1940. Part of the material of this book is available in English in the article "The Nature of Economic Regions," *Southern Economic Journal*, V (July 1938), pp. 71–78, and in a review article by W. F. Stolper, *American Economic Review*, XXXIII (September 1943), pp. 626–36. Also see Lösch's article, "Beiträge zur Standortstheorie," *Schmollers Jahrbuch* LXII (1938), pp. 329–35.

through a set of elementary equations a highly simplified static model of a space-economy operating *under conditions of monopolistic competition*. To appreciate fully this model, one must understand Lösch's concept of the market, by means of which space is introduced into the problem, and which represents his other major contribution to location theory. What is the market area? How is it bounded? These questions are fundamental.

Lösch postulates the following: a broad, homogeneous plain with uniform transport features in all directions and with an even scatter of industrial raw materials in sufficient quantity for production; a uniform distribution of agricultural population with a uniform set of tastes and preferences, each homestead at the start being self-sufficient; technical knowledge disseminated throughout the plain and production opportunities available to all. In all other respects, too, extra-economic forces are excluded. If in this situation an individual finds it profitable (owing to the economies of large-scale production as opposed to the handicap of transport cost) to produce a commodity over and above the needs of his homestead, his market area would assume a circular form. However, if one farmer finds it profitable to produce over and above his needs, so will others, and the force of competition, by eliminating all excess profits, not only will contract the market area of the original producer, but also will transform the circular shape of market area into a hexagon. The hexagon is the ideal economic form of market area, it is maintained. Firstly, a net of hexagonal market forms will exhaust (completely cover) any area under consideration, whereas circular ones will leave empty, unutilized corners, as is readily seen from a graphic presentation. Secondly, of all the polygons (hexagon, square and triangle) which will exhaust a given area, the hexagon deviates least from the circle form, and, in consequence, minimizes the transport expenditures in supplying a given demand, or, expressed differently, maximizes the demand of the population of a given area.[5]

For each commodity, then, the plain is dissected into a honeycomb (a net of hexagons) of market areas. Lösch next groups these honeycombs according to the size of their respective market units; and, in a manner consistent with the established criterion of minimum transport effort, he orders the resulting nets about a common, central production point to obtain his system of nets.[6]

We are now in a position to reproduce Lösch's scheme of general equilibrium. At the start Lösch attacks the problem of the location

5. *Die räumliche Ordnung* . . . , *op. cit.*, pp. 65–72.
6. *Ibid.*, pp. 73–84.

of the production of industrial goods alone. The same hypotheses which were basic to the above determination of market areas are retained for the general analysis. Table I presents the symbols of spatial arrangement. The position in the plain, of each production place of each commodity, is designated by a set of $x$, $y$ coordinates; the boundary of the market area of each production place is described by a set of equations, each equation being represented by a corresponding Greek symbol in Table I.

Lösch puts forth the following, either as given or unknown:

A. Given:

$G$ = total surface area
$m$ = number of products
$r$ = freight rate
$\sigma$ = rural population per sq. kilometer
$\sigma_q^m$ = population of the city $p_q^m$
$d^m = f^m(\pi)$ = individual demand for the good $m$
$\pi_q^m = \varphi^m(Dq)$ = the factory price $\Big\}$ of the good $m$ at place $q$ as function of
$K_q^m = \chi^m(Dq)$ = the average production cost $\Big\}$ the total demand $D_q^m = \Psi(f^m, x_q^m, y_q^m, \alpha_q^m, \beta_q^m \ldots \epsilon_q^m, \sigma, \sigma_q^m \ldots)$.

B. To be sought:                                                        *No. of Unknowns*

1.  $\pi_q^m$              = factory price of the                        $n$
                             good $m$ at location $p_q^m$

2.  $G_q^m$               = market area of the location                  $n$
                             $p_q^m$ in sq. kilometers

3.  $q^m$                 = the number of towns that                     $m$
                             produce good $m$

4.  $x_q^m, y_q^m$        = coordinates of the location $p_q^m$          $2n$

5.  $\alpha_q^m, \beta_q^m \ldots \epsilon_q^m$  = equations of the boundary of the          $N$
                             market area of $p_q^m$

                                                            Total: $4n + m + N$

Corresponding to the list of unknowns, Lösch presents in Table II a set of equilibrium equations. The first condition for equilibrium which must be fulfilled is that no abnormal profit should exist. The cost of each commodity produced must equal its factory price. But further, the assumption of free entry into any form of production guarantees the second condition of equilibrium, namely the Chamberlinian tangency solution. (Owing to the spatial dimensions of the market the demand curve for the good produced at any place must necessarily slope negatively.) The changes in average price and average cost ensuing from an infinitesimal change in the space dimension of the market must be equal: in other words, each individual producer's market area must be the minimum which is economically possible. And, thirdly, the whole plain under consideration must be

TABLE I. SYMBOLS OF SPATIAL ARRANGEMENT

| Product No. | Production Places | | Market Boundaries | |
|---|---|---|---|---|
| | Position | No. | Abbreviations of their equations | No. |
| 1 | $p_1^1(x_1^1 y_1^1); p_2^1 \ldots$ | $a$ | $\alpha_1^1, \beta_1^1 \ldots \epsilon_1^1; \alpha_2^1, \beta_2^1 \ldots$ | $A$ |
| 2 | $p_1^2(x_1^2 y_1^2); p_2^2 \ldots$ | $b$ | $\alpha_1^2, \beta_1^2 \ldots n_1^2; \alpha_2^2, \beta_2^2 \ldots$ | $B$ |
| . | | . | | . |
| $m$ | $p_1^m(x_1^m y_1^m); p_2^m \ldots$ | $q$ | $\alpha_1^m, \beta_1^m, \ldots p_1^m, \alpha_2^m, \beta_2^m \ldots$ | $Q$ |
| $m$ | (total) | $n$ | (total) | $N$ |

$$= \Sigma a + b + \ldots + q$$

$$= \frac{\Sigma A + B + \ldots + Q}{2}$$

TABLE II. EQUILIBRIUM SYSTEM I

| Condition | Fulfilling Equation | No. of Equations |
|---|---|---|
| 1. No abnormal profits | $\varphi^m(D_q) = \chi^m(D_q)$ | $n$ |
| 2. Area as small as possible | $\dfrac{\partial \pi_q^m}{\partial G_q^m} = \dfrac{\partial K_q^m}{\partial G_q^m}$ | $n$ |
| 3. Total area utilized | $\Sigma G_1^m + G_2^m = \ldots + G_q^m = G$ | $m$ |
| 4. Maximum demand | $\dfrac{\partial D_q^m}{\partial x_q^m} = 0; \quad \dfrac{\partial D_q^m}{\partial y_q^m} = 0$ | $2m$ |
| 5. Boundary indifference lines | For any point $x, y$ on boundary equation $\alpha$: $$\pi_q^m + r_q^m \sqrt{(x - x_q^m)^2 + (y - y_q^m)^2}$$ $$= \pi_{q-1}^m + r_{q-1}^m \sqrt{(x - x_{q-1}^m)^2 + (y - y_{q-1}^m)^2}$$ | $N$ |

Maximum number of producers { 1, 2, 3 }

exhausted by the various market areas. These three conditions insure the maximum number of independent producers.

The fourth condition of equilibrium stipulates that the individual producer occupy a spatial position which maximizes his demand; as a result he will not find it desirable to change his location either in the $x$ or $y$ direction. The fifth condition requires that any consumer on any boundary line be indifferent as to the possible production sources from which he can obtain a given commodity at a minimum cost.

Since, in toto, the number of fulfilling equations of Table II equals the number of unknowns, the system of spatial economy is determinate; the unknowns can be derived.[7]

In a manner analogous to the above, states Lösch, the location of the production of agricultural goods can be fairly easily analyzed. In a similar manner, too, the reverse propositions, which concern the conditions for the best location of industrial and agricultural places in their capacity as consumption centers, can be attacked. Unfortunately, however, the optimal location for production does not necessarily coincide with the optimal location for consumption.[8]

This is the way in which Lösch spins his web of general equilibrium. Although his model is highly simplified, and "over-assumed" or "over-determined," although the elements of interdependence are minimized and his approach does not comprehend the space economy as a whole, but as consisting of several major sectors, we have here for the first time an attempt to encompass general spatial relations in a set of equations. And through eschewing the assumption of pure competition and postulating monopolistic competition in its stead, Lösch goes far in meeting Palander's objections to spatial general equilibrium analysis.

One need not, however, proceed, as does Lösch, in deriving a set of equations. Lösch assigns a set of spatial coordinates to each producer and consumer. This permits, in a sense, a geographic description of a space-economy. But his presentation would become exceedingly complex if one were to relax his simple uniformity assumptions which are basic to his model — if one were to allow inequality in raw material, labor and capital resources, an uneven and discontinuous distribution of population, and all other types of local differences. To introduce inequality in the spatial pattern of inputs alone is a very difficult task.

7. *Ibid.*, pp. 57–60. Tables I and II and the lists of given and unknown conditions are for the most part literal translations.
8. Lösch does not develop specifically these additional sets of equations.

From a *functional* standpoint — one that is relevant to the incessant struggle within the economy as a whole, as well as within its various parts, to attain the correct set of substitution points with respect to inputs, outputs, outlays and revenues — Lösch's model is anemic. It is much more meaningful to design a set of equations depicting general equilibrium in terms of input-output relations and price-cost relations, including therein distance inputs (and if possible local price-cost variations) in order to give explicit recognition to the factor of space. This latter model, constructed without reference to sets of spatial coordinates, would be much more able to cope with further possible theoretical developments, for example with the structural, dynamic developments visualized by Weigmann; although, to be sure, any spatial description of the order of Lösch is desirable if it does not impose restrictions upon the basic operations of the model.[9]

With respect to input-output relations, the Leontief technique, within the limits to substitution imposed by its assumptions, offers a very powerful tool of analysis.[1] By incorporating distance inputs into the Leontief model and by further detailing its tables through decomposing the nation into regions, the regions into states, the states into smaller geographic units, and so on, one can give an increasing amount of play to the spatial substitution operations. Such a model can thus reflect to a large extent the interindustrial relationships of the space-economy.

Apropos of price-cost relations, it cannot be too strongly emphasized that the theories of space-economy and of monopolistic competition (broadly conceived[2]) are inextricably bound together. The noteworthy contribution of Chamberlin in developing techniques for spatial analysis has not been treated specifically since it has been largely digested by Lösch and Palander. Progress along Chamberlinian lines, however, is a *sine qua non* for developing further the theory of the space-economy, especially in its welfare aspects.[3] Triffin has already built upon Chamberlin's structure, setting the monopolistic

9. For other relevant criticism of Lösch, see W. F. Stolper, *op. cit.* and Hans Ritschl, "Aufgabe und Methode der Standortslehre," *Weltwirtschaftliches Archiv*, LIII (1941), pp. 115–25.

1. W. W. Leontief, *The Structure of American Economy, 1919–1929* (Cambridge, Mass., 1941).

2. To include oligopoly with or without product differentiation. See E. H. Chamberlin, "Some Final Comments," *The Review of Economics and Statistics*, XXXI (May 1949), pp. 123–4.

3. For most situations of the space-economy, it is quite meaningless to apply the norms of pure competition. Also see Chamberlin, *Theory of Monopolistic Competition*, 5th or 6th eds., pp. 213–15.

competition techniques in a general equilibrium framework.[4] Triffin's interdependence analysis, in many places explicitly cloaked in substitution terms, is not unlike Predöhl's substitution technique (although, to be sure, Triffin hardly thinks in terms of spatial or location relations). In this sense, then, a generalized theory of monopolistic competition, broadly defined, and including the physical production (input-output) problem in its spatial setting, can be conceived as synonymous with our general theory of location and space-economy.

## VI

One final matter should be discussed, namely, the interrelation of trade theory and the general theory of location and space-economy. In 1911 Weber pointed out that classical trade theory ignored entirely the transport cost involved in traversing space.[5] He particularly criticized the classicists for overlooking the large portion of internationally distributed industry which is transport-oriented and which seeks the minimum transport cost point with respect to raw materials and markets, and for attributing to international division of labor and capital the international distribution of transport-oriented industry.

Furlan, Engländer, Ritschl, Weigmann and others have stressed the interrelation of trade and location theories, but not until the appearance of Ohlin's *Interregional and International Trade*[6] do we have a serious attempt to integrate the two. As one of his objectives Ohlin purports:

"to demonstrate that the theory of international trade is only part of a general localization theory, wherein the space aspects of pricing are taken into full

4. R. Triffin, *Monopolistic Competition and General Equilibrium Theory* (Cambridge, Mass., 1940). Although the monopolistic competition of Chamberlin seems to be more than a particular equilibrium theory (certainly it is at least a quite broad particular equilibrium theory, for it embraces the problems both of *individual* equilibrium and of *equilibrium* for an elastically defined *group*), nonetheless it is not a general equilibrium theory in the full meaning of the term (see Triffin, pp. 8–9, 54, 67 and elsewhere). Triffin's contribution consists of extending the scope of monopolistic competition to encompass the complex net of competitive interrelationships throughout the entire economic collectivity. In doing this, Triffin discards the concepts both of an industry, and of a group of firms. Rather, he emphasizes the individual firm (or more strictly, the maximizing unit) and the various coefficients of interdependence between any given firm and each of all the other firms in the economy, both with respect to factors and to products.

5. "Die Standortslehre und die Handelspolitik," *Archiv für Sozialwissenschaft und Sozialpolitik*, XXXII (May 1911), pp. 667–88.

6. (Cambridge, Mass., 1933.) Some of his other relevant works are: *Handelns teori* (Stockholm, 1924) and "Some Aspects of the Theory of Rent: von Thünen vs. Ricardo," *op. cit.*

account, and to frame certain fundamentals of such a theory as a background for a theory of international trade, wherein the influence of local differences in the supply of factors of production and transportation costs within each country is duly considered."[7]

Ohlin plants his objectives within the framework of a mutual-interdependence theory of pricing, the latter to be expanded to enfold the multitude of markets and local price variations which ensue from the varying spatial immobilities and indivisibilities of goods and factors. Thus his general localization theory would determine simultaneously prices, markets, location of industry, commerce and agriculture, spatial distribution of factors and commodities and other economic magnitudes.

It would seem logical that Ohlin should first develop a general localization theory. Then, by focusing upon certain forms of immobilities of factors and goods (consideration of other relations set aside for the time being) he could develop at length his theory of interregional and international trade as a special case. Unfortunately, Ohlin adopts an entirely different procedure and as a consequence has to employ a somewhat unique casuistry. Parts I and II of his book are devoted to the theory of interregional trade and a simplified version of international trade, respectively. These parts, however, are constructed upon an unrealistic set of hypotheses. The region is defined as that area within which there is perfect mobility of factors. Between regions factors are considered perfectly immobile. And all impediments to movement of commodities are assumed away. In Part III he attempts to approach reality through the successive introduction of the following: (1) interregional costs of transfer of commodities; (2) interregional factor movements; (3) interior costs of transfer and factor movement; (4) local differences in labor and capital supply. The inclusion of the last two items represents an effort to subject the theory of interregional trade to a broadening process and thus convert it into a general localization theory.

Ohlin at most achieves a weak (and only verbal) general localization theory. He does not attain the total systematic analysis which characterizes his interregional trade theory. His treatment of locational forces (Chapters X–XII, inclusive) is quite sketchy and flimsy. At the start, a modified version of Thünen's isolated state is applied to industrial production within a district, "the frontiers of which are not described." Ohlin imagines at first that his district possesses "uniform transport features" throughout its area, and that

7. *Op. cit.*, p. vii.

within it the factors of labor and capital are perfectly mobile. At the center lies a strategic natural resource, perhaps coal or iron ore deposits. The surrounding zones of cultivation of various agricultural products, rent of land and prices of commodities can be determined only through a mutual-interdependence system. Next, the general approach on the whole is abandoned in favor of a step-by-step analysis commencing in typical Weberian style. The localization of manufacturing, of raw material production, of consumers' markets, local differences in transport resources and facilities, economies of large-scale operation and concentration, local differences in capital and labor supply are successively considered. In the end, however, Ohlin returns to a general interdependence setting in depicting the relations of the various economic forces.

This singular approach has turned out to be misleading to many. One sympathetic critic maintains that Ohlin does not successfully bridge the gap between interregional trade theory and general localization theory, and thus does not achieve a unified theory; for the district whose total localization is supposed to be explained by a general theory does not necessarily have to possess the same mobility characteristics as the region, which is the unit of study for interregional trade.[8] To the extent that the exposition of the total analysis for the district is deficient (and of this there is no question), the gap is not bridged, but the district itself can be conceived as boundless, or more realistically, as synonymous with the world (which Ohlin does not explicitly do). A satisfactory, exhaustive, total analysis for the district would then describe all economic relations within the world and explain all manner of trade. One could then deduce interregional trade analysis (no matter on what basis the region and subregions are delineated) by singling out from among the complex of relations those of relevance.[9]

8. Tord Palander, *op. cit.*, pp. 266–67. Elsewhere (pp. 262–64) Palander summarizes Ohlin's earlier study, *Handelns teori*. Here an attempt is made to extend interregional trade to a theory of interlocal trade (thus accounting for local differences in factor supply) through subdividing the region (within which perfect mobility of factors reigns) until the sub-regions become identical with the localities themselves. Obviously this technique is inadequate; it assumes away the location problem, for at the start the basis of interregional trade is presumed to be the different relative scarcities of productive factors among regions.

9. Lösch ("Beiträge zur Standortstheorie," *op. cit.*, p. 331) has also charged Ohlin with lacking a clear answer to the location problem within his regions, and with a failure to perceive labor distribution as a result of economic activity between men, not between regions. The latter accusation falls, however, with the definition of the district as the world, or as an area greater in extent than, and inclusive of, the region. ·

On the other hand, one may perhaps with some justification raise objections to the liberal and generous interpretation given herein to Ohlin's reasoning.

However, it is not necessary at all to view trade theory as narrowly as Ohlin does. It is true that international trade theory historically and as it exists today in such standard works as Viner's[1] and Haberler's,[2] does correspond to Ohlin's conception of it. It is still subject to Weber's criticism: it does not incorporate transport-oriented industry into its analytical framework, and is thus inadequate for determining policy. Nonetheless, one can view trade theory and the general theory of location and space-economy as synonymous. For (1) location cannot be explained without at the same time accounting for trade and (2) trade cannot be explained without the simultaneous determination of locations. Once we recognize this it is futile to argue whether trade theory is or is not a special form of general location theory.[3] But since it is really a matter of indifference how we define trade theory and relate it to location theory, it may be wise to adhere, as Ohlin does, to the traditional scopes of these theories. Trade theory may be regarded as a part of a general theory of location and space-economy, and synonymous with that partial analysis which proceeds from the assumption of a given locational framework of economic activities.

## VII

In summary, the general theory of location and space-economy is conceived as embracing the total spatial array of economic activities, with attention paid to the geographic distribution of inputs and outputs and the geographic variations in prices and costs. Modern general equilibrium theory is a special case of this theory, in which transport costs are taken as zero and all inputs and outputs are viewed as perfectly mobile; international trade theory, in its traditional scope, is also a special case of this theory. One proceeds from the latter to the former by assuming a given locational structure of economic activities, by erecting appropriate barriers within the world economy to correspond to the boundaries of nations, and so forth. However, it is important to bear in mind that the distinction between trade theory and the general theory of location and space-economy is one of definition only. Trade theory can be redefined to be synonymous with the general theory of location and space-economy. And in a sense, too, because of the monopoly elements which are almost

1. *Studies in the Theory of International Trade* (New York, 1937).
2. *The Theory of International Trade* (London, 1936).
3. Thus Viner's cynical remark about Ohlin's dictum that the theory of international trade is nothing but international location theory is really unnecessary and indicates either Viner's confusion or his failure to appreciate the scope of location theory (*op. cit.*, p. 468 n.).

invariably present in spatial relations, a broadly defined general theory of monopolistic competition can be conceived as identical with the general theory of location and space-economy.

The substitution principle provides the best analytical tool for developing this general theory. However, Predöhl's use of this tool must be modified and extended to embrace various substitution relations between distance inputs, local outlays and local revenues. The formulation of these relations in terms of a system of mathematical equations ought first to embrace the concept of distance inputs, and later, if possible, Lösch's sets of spatial coordinates.

The evolutionary approach of Weber and others, and especially the writings of Weigmann who conceives the space-economy as a rhythmic-moving Gestalt whole with a basic structural core of land and labor markets, will be very helpful in suggesting lines along which this general theory may be nurtured to embody dynamic relations.

<div align="right">WALTER ISARD.</div>

HARVARD UNIVERSITY

# ENGINEERING PRODUCTION FUNCTIONS[1]

## SUMMARY

The problem, 507. — I. The scope of engineering analysis, 508. — II. Use of engineering data in economic analysis, 510. — III. Production-function analysis of pipeline transportation, 514; the technological relations, 514; cost functions, 517; conditions of equilibrium, 519. — IV. The use of engineering production functions: results of the present study, 526; further applications, 529; conclusions, 531.

The conventional division between economics and engineering which has been assumed by economists is a purely arbitrary separation. It is usually indicated by saying that engineering is concerned with alternative methods of achieving a given productive goal, while economics takes these methods as given, and analyzes the interrelations among productive ends and the factors used to achieve them. The division of labor in the analysis of production which is suggested by this separation has obvious advantages. In practice, however, the difference between the engineering and the economic approach has limited the practical application of economic theory because the results of engineering studies are rarely formulated in such a way that they can be used by economists.

Because of this inability to use engineering data as a basis for economic reasoning, a great discrepancy exists between the theoretical analysis of the problems of production and the empirical studies which have been made. Economic theory is based on a hypothetical production function which describes a productive process by specifying the greatest amount of a given product which can be produced by any combination of productive agents or inputs. Only in the field of agriculture, however, have economists attempted to determine the shape of these production functions empirically. Industry studies have generally used statistically determined cost curves. Since these curves are based of necessity upon productive combinations which it has proved feasible for entrepreneurs to try out, they cannot usually tell us much about the broader range of productive possibilities which have been explored experimentally but not adopted commercially. The lack of this information is a great handicap in many types of

economic discussion.  It prevents quantitative analysis of the possibilities of substituting one factor for another, since statistical data only pertain to historical observations in which effects of technological change and price variation are usually inseparable.  The lack of information on substitution restricts the factual analysis of production to the short period for which productive units may be assumed to be constant; only then can statistically derived cost curves be a valid approximation to current conditions.

The purpose of the present paper is to suggest a method by which engineering data may sometimes be used to approximate the production functions of economic theory in the industrial field, and to derive one such function.[2]  While the industry studied is a particularly simple one and therefore not typical of all the problems to be encountered, an actual example of the use of engineering relations may be of value in clarifying the problems which must be met in bridging the gap between engineering and economics.

## I. THE SCOPE OF ENGINEERING ANALYSIS

Before suggesting a way of using engineering data in economic analysis, we must consider the problems which the engineer himself is trying to solve.  Since his initial aim is to discover all feasible ways of making a given product or performing a given service, his first concern is not with particular inputs but with the nature of the chemical or physical transformations which are involved in the productive process.  He breaks down the process of production into convenient units whose performance he attempts to describe by formulae based on the laws of physics and chemistry.  Since an elemental analysis in terms of the properties of each piece of equipment is often impractical, the engineer must usually resort to testing various sizes and combinations of equipment to determine the effect of such variables as size, speed, temperature, etc. upon total performance.  The probable output of a whole industrial plant is estimated by a series of calculations of this sort.  One basic difference between engineering analysis and economic analysis, then, is the units which are considered fundamental.  While

2. Although the results of engineering studies are sometimes used by economists to obtain cost curves, to my knowledge experimental data has been used systematically only to derive production functions in the field of agriculture.  The analysis of air transport by M. Louis Bréguet (summarized by E. H. Phelps Brown in *Econometrica*, 1936, pp. 249–259) uses a technique to derive a cost curve based on engineering experience which is in many ways similar to that which will be used in this paper.  Professor Stigler (*American Economic Review*, March 1940, p. 402) and others have suggested that an engineering approach would have advantages in the derivation of production functions and cost curves, but the suggestion does not seem to have been taken up by economists.

the economist deals with plants or firms or industries, the engineer must deal primarily with separate physical processes.

The difference between the economist's problem and that of the engineer is shown most clearly by the variables which they find it convenient to use. In describing the process of production, both must deal with the quantity, quality, and price of all feasible inputs which may produce a given output. The engineer is interested first in the selection of inputs and second with the quantity of each which will be required. He normally treats prices as parameters which will either be held constant throughout the analysis or at most take only a few possible values. The economist on the other hand is interested primarily in the effect of varying prices upon productive combinations. He, therefore, treats each qualitative variation in an input as if it were a separate input in order that each one will have an individual price. This procedure is useful so long as the discussion remains on an abstract level, but it is very cumbersome to apply in analyzing an actual productive process.[3]

Since the engineer's primary aim in the first stages of analysis is to determine what inputs may be combined to produce a given result, he describes the process in terms of the properties of the inputs. For example, to determine the input requirement of an unspecified metal, he may state the hardness, tensile strength, and resistance to shear necessary to produce a given result. At this point cost calculation using the assumed prices is necessary to choose the metal which will prove most economical. The quality and type of the inputs, which the economist assumes as parameters, are the end result of this phase of engineering analysis. Since the design of an industrial plant consists in a long series of such calculations, the engineer does not usually consider a wide range of price possibilities but merely adopts the combination of inputs which produces the least cost with the assumed prices. The engineer's design calculations are therefore usually limited in their applicability to the few price combinations which have been considered. It is not feasible for him to maintain the degree of generality which the economist prefers in reserving all cost calculations until all possible physical combinations of inputs have been considered. This method would greatly increase the work of designing even a simple plant; therefore the engineer makes a series of least-cost calculations as he goes along.

3. The dimension which the economist calls "quantity" might be defined as that variable to which a price is attached. The necessity for this correspondence between quantity and price leads him to formulate the production function in a discontinuous way. The engineer seeks to avoid such a discontinuity.

If the economist wishes to use engineering data to construct a production function, he must go back to the intermediate stage in engineering calculations at which the possibilities of using various types[4] of inputs are considered. These data are found in engineering textbooks but not in the published results of engineering studies. In order to use it conveniently, the economist must abandon his convention of using one-dimensional inputs and use multi-dimensional inputs as the engineer does. The advantages of this procedure will be demonstrated below.

The usefulness of engineering data to economists varies from industry to industry. Engineering as commonly conceived concerns itself principally with the performance of machines rather than of men. Consequently the accuracy of a production function based on engineering data alone will vary inversely with the variability of the labor input. The accuracy of an engineering production function will likewise vary considerably among industries according to the development of engineering science in the particular case. A chemical reaction may be described more exactly than the design of a machine, for example. The effect of varying one input separately from the other inputs required may sometimes be accurately determined from engineering considerations, however, even where calculations of the total output have a considerable margin of error.

The simplest case for the construction of a production function from engineering considerations alone would be one in which labor can be treated as a joint factor with some other input, in which engineering science is well developed, and in which the technical characteristics of one or a few principal processes are a determining factor in the cost structure of the plant or firm. The industries which seem to come closest to this ideal are the chemical processes, the refining of raw materials, and other standardized, automatic techniques.

## II. Use of Engineering Data in Economic Analysis

### A. *Selection of Independent Variables*

Let us first define precisely three classes of variables which will be used in this paper. The familiar economic categories of land, labor, capital, materials etc. will be called economic factors of production. The purchasable commodities which make up these general categories will be called physical inputs. It is these which appear in the economist's production function. The third category of variables

---

4. "Types" of inputs are to be distinguished by differences in one or more physical dimensions other than quantity. For example, copper, aluminum, and steel might be alternative types of a metal input.

consists of the physical properties (or dimensions) of the inputs or of the process itself, which will be called engineering variables.

It is proposed to rewrite the production function of economic theory in terms of engineering variables. For this procedure to be valid, it is necessary that the quantity and cost of the physical inputs be determined by the variables chosen. Several examples may make this proposal clearer. In the process of steam generation, the output of steam per hour is a function of the rate of flow of water, the temperature, and the dimensions of the boiler. These in turn determine the pressure in the boiler, the stress in the steel, and hence the type and amount of steel required. The amount of fuel needed can be analyzed in the same terms. To determine the most economical type of fuel, a calculation must be made for each alternative source. This is typical of all cases where there are several alternative processes. The final production function may involve different processes over different ranges of output. For example, if the principal element in a process is an automatic machine, the relevant engineering variables may be its speed, size, continuity of operation, etc. The analysis of machine processes is likely to be more complicated than that of fluid operations, and less exact.

It is difficult to specify *a priori* any rules for the selection of independent variables other than the requirement that they must uniquely determine both quantities and types of inputs for a given output and given costs. The aim is to try to reduce the actual design calculations of the engineers to a manageable form while retaining a maximum of information as to the possibility of substituting one input for another.

B. *The Engineering Production Function*

A function which fulfills the above conditions may be called an engineering production function. By a simple transformation it may be used in the same way as the usual relation postulated by economists in terms of physical inputs or of economic factors.

The conditions assumed above may be expressed as follows:

$$u_i = u_i (v_1 \ldots v_n) \qquad (i = 1, \ldots, n) \qquad (1)$$

$$p_i = p_i (v_1 \ldots v_n) \qquad (2)$$

where: $u_i$ is the quantity of each physical input,

$p_i$ is the price of $u_i$ per unit,

the $v$'s are the engineering variables.

The production function assumed in economic theory may be written:

$$X = f(u_1, \ldots u_m) \qquad (3)$$

Where $X$ is the output per unit of time (or the capacity of the process).

The corresponding engineering function is then:

$$X = \phi(v_1, \ldots v_n) \tag{4}$$

since the $u$'s in equation (3) may all be expressed in terms of $v_1 \ldots v_n$.

The cost function may be similarly transformed. The total cost is:

$$C = \sum_1^m u_i \, p_i \tag{5}$$

This may be rewritten:

$$C = \pi(v_1, \ldots v_n) \tag{6}$$

since both quantities and costs are functions of the engineering variables.

The usual mathematical analysis of the theory of cost and production can be rewritten in terms of equations (4) and (6), but this is unnecessary because no change has been made in the form of the equations. The cost per unit of the engineering variables is not constant for changes in quantity, but this difference does not greatly complicate the problem. Only the first steps in the general analysis will be repeated here.

### C. *Conditions of Equilibrium*

Following Professor Samuelson,[5] we may write the necessary conditions for the minimum cost of any given output, $\overline{X}$ as follows, using the method of the Lagrangean multiplier:[6]

Let

$$Z = \pi(v_1, \ldots v_n) - \lambda[\phi(v_1, \ldots v_n) - \overline{X}]$$

For a proper relative minimum,

$$\frac{\partial Z}{\partial v_i} = \pi_i' - \lambda \phi_i' = 0$$

from which

$$\frac{1}{\lambda} = \frac{\phi_1'}{\pi_1'} = \frac{\phi_2'}{\pi_2'} = \cdots = \frac{\phi_n'}{\pi_n'} \tag{7}$$

The $n$ equations in (7) plus equation (4) determine the values of the $n$ engineering variables (and of $\lambda$) which will produce a given output, $\overline{X}$, at least cost. As in the case of physical inputs, the marginal

5. All of the mathematical analysis of Samuelson can be applied to engineering production functions by substituting my equations (4) and (6) for his eq. (1) and (3) in Ch. IV, *Foundations of Economic Analysis.*

6. Any limitational relations among the variables in the production function would necessitate breaking it down into several equations with an equal number of multipliers.

productivities, $\phi$, must be proportional to the marginal costs for each variable. The secondary conditions for a proper minimum are fulfilled if the isoquants in terms of the $v$'s are convex to the origin in all directions.[7] When factor costs are not taken to be constant, these conditions are the same as those for physical inputs. The quantities and prices of the physical inputs are then determined by substituting the equilibrium values of the $v$'s in equations (1) and (2).

One important use of production functions is to determine the change in the equilibrium values of the inputs resulting from a change in one of the cost elements. These cost elements enter equation (6) as parameters, but any one of them may be treated as a variable. Examples of the variation in the equilibrium values resulting from a change in particular prices will be given in the next section. For a general mathematical treatment, the reader is referred to Professor Samuelson's analysis.[8]

The relationship between engineering production functions and those which the economist is accustomed to use should now be clear. The engineering function describes each input in terms of one or more of its properties, which may vary independently within limits. The corresponding function in terms of one-dimensional inputs would require as many variables as the possible types of each input. Use of multi-dimensional inputs makes the mathematical analysis much simpler.

Although it is not in general possible to transform an engineering production function into one involving one-dimensional inputs without greatly multiplying the number of variables, it is always theoretically possible to write it in terms of any general categories such as capital, labor, materials, etc., which the economist chooses. To take a simple two-variable case, each physical input implies a certain amount of capital and current inputs. The amounts of capital and current inputs can therefore be specified in terms of the engineering variables. Since different combinations of the $v$'s may require the same amount of capital, the production function can only be defined by minimizing the amount of current inputs for each quantity of capital. Mathematically, this is a constrained minimum problem with two subsidiary conditions: output and capital constant. For $n$ factors, there will be $n$ such side relations.[9] Hence, even a very simple engineering production function may result in a relation

---

7. This condition holds unless the isocost surfaces are convex to a greater degree in some region.

8. *Op. cit.*, pp. 63–69.

9. For the general solution, see Samuelson, *op. cit.*, p. 362.

between capital, labor, and materials which is difficult to deal with mathematically. Since the economic relationship is implicitly defined by the engineering production function, it will usually be more convenient to deal with the latter throughout the analysis. This procedure will be followed in the example below.

The use of multi-dimensional products has already been suggested in the field of consumption, but there the analysis is hampered by the fact that neither cost nor consumer preferences show a consistent variation with the various properties suggested.[1]  In the field of production, however, these obstacles are absent because consistent functional relationships exist in general between the dimensions of the inputs and both output and cost.

III PRODUCTION-FUNCTION ANALYSIS OF PIPELINE TRANSPORTATION[2]

One of the simplest processes known to the writer which serves as an example of the suggested technique is that of movement of fluids by pipe. Since this is the entire productive function of several industries, it has considerable economic importance as well. Although all formulae for fluid flow are essentially similar, the case of natural gas transportation has been chosen because the compressibility of gas makes the range of variation in production possibilities more extensive. The economies of scale which will be demonstrated by this example are typical of all pipeline flow and to some extent of all processes involving containers. Fluid flow recommends itself for this type of analysis because the formulae required can be determined quite accurately in the laboratory and modified through experience with larger-scale operations. There is only one feasible technique,[3] and the production function is continuous.

A. *The Technological Relations*

Natural gas is transported in volumes up to half a billion cubic feet per day through pipelines ranging in diameter to over thirty inches, and in length to over 1500 miles. The amount of gas trans-

1. The difficulties of quantitative treatment of "product variation" are shown by Professor Chamberlin in *The Theory of Monopolistic Competition*, Chapter V. See also H. Brems, "The Interdependence of Quality Variations, Selling Effort and Price," this *Journal*, May 1948, p. 419.
2. I am indebted to Mr. Reginald Burdick, vice-president of the Southern Natural Gas Company and to Mr. Ross Holmes for information on the design and operation of natural gas pipelines.
3. I.e. only pipeline transportation is economically feasible regardless of the variation in prices and its productive possibilities can be described by a single function.

mitted by a pipe depends upon its diameter, the pressure of the gas, and the pressure drop along the line. Hence capacity may be increased by either increasing the diameter, the pipe thickness, or the pumping capacity. An empirical relationship among three engineering variables which has been commonly used in the design of pipelines is given by Weymouth's formula:[4]

$$X = K \, D^{8/3} \, \sqrt{P_1^2 - P_2^2} \tag{8}$$

where: $P_1$ is the initial pressure in the pipe (in pounds per square inch absolute),

$P_2$ is the outlet pressure,

$D$ is the inside diameter of the pipe in inches,

$X$ is the capacity in million cubic feet of gas per day,

$K$ is a parameter which depends upon the conditions of measurement, the specific gravity of the gas, its flowing temperature, and the length of the pipeline between compressor stations.

This relation may be rewritten as a function of the initial pressure and the pressure drop through the section:

$$X = K \, D^{8/3} \, P_1 \, \sqrt{1 - (\tfrac{P_2}{P_1})^2} \tag{9}$$

Since all inputs are proportional to the length of the line, we may choose 100 miles as a convenient unit of analysis; the results will apply to any length.

The production relation for gas flow has been written in terms of pressures and the diameter of the pipe. The pressures must be expressed in terms of the properties of the physical inputs, pipe and the compressors, in order to satisfy the conditions for an engineering production function. This transformation is effected by making use of the following properties of pipe and pumping engines:

(a) Working pressure of a pipe depends upon its thickness, diameter, and allowable maximum stress, as follows:

$$P_1 = \frac{2ST}{D} \tag{10}$$

where: $S$ = working stress (pounds per sq. inch)

$T$ = thickness (inches)

(b) The horsepower of a compressor is a function of the com-

4. See Pacific Gas Assn., *Gas Engineer's Handbook*, 1934, p. 727ff. Weymouth's formula is derived from purely thermodynamic reasoning except for the introduction of an empirical friction factor. It ignores the deviations from the perfect gas law at high pressures and hence is inaccurate above about 500 P.S.I.

pression ratio between outlet and intake and the quantity of gas pumped:[5]

$$H = (28.75R - 13.9)X \qquad (11)$$

approximately, over the practical range of $R$.

where:          $R = P_1/P_2 =$ compression ratio
                $H =$ horsepower

(c) The weight of steel pipe depends on its diameter and wall thickness:          $W = 28.2(D + T)\ T$

or approximately          $= 28.2\ DT \qquad (12)$

Formulae (10) and (11) enable us to rewrite equation (9) in a more useful form, using the stress as a parameter.

$$X = K_1\ D^{5/3}\ T\ \sqrt{1 - 1/R^2} \qquad (13)$$

where $K_1 = 2.623$ for a value of $S = 15,000$ p.s.i. and a length of 100 miles.

Equations (11) and (12) are the required relations between the physical inputs and the engineering variables specified in (1) above. The input of pipe may be described in terms of any two of the three variables $D,\ T,$ and $W$. Equation (11) is actually a separate production function for the process of compression, which is used to translate the compression ratio $R$ into the horsepower required. All costs of compression are assumed to depend on the horsepower per station.

The principal problems in designing a natural gas line are the determination of the thickness[6] and the diameter of the pipe and the size and spacing of compressor stations.[7] Variation in station spacing is an alternative for varying the horsepower per station, but it is taken as a constant here because its effect on the final results is small.[8]

5. The true function for adiabatic compression is:
$$H = A(R^{.213} - k)X$$
where $k$ depends on the efficiency of the particular compressor. A value of $k = .96$ is used here. See H. C. Lehn, "An Analysis of Gas Pipeline Economics," *American Society Mechanical Engineers Transactions*, July 1943, p. 446.

6. In the following solution, the fact that pipe thickness has a minimum practical value which varies with diameter is ignored for the sake of simplicity. In practice, a theoretical value of $T$ of less than the minimum (between .18 and .30 inches) would indicate the use of the smallest practical value and a corresponding revision of the cost curves. The resulting difference in total cost is slight.

7. When manufactured gas is transported, the heat content of the gas may also be varied. The effect of this variable upon transport costs is shown by M. A. Van der Hende, "Estimation de la Distance Economique du Transport du Gaz," *World Power Conference, Fuel Economy Section*, 1947.

8. In most engineering analyses of this problem, the horsepower per unit of capacity is solved for its optimum value and then held constant while the spacing

In addition to the required quantity of gas to be transported, the given conditions include the estimated load factor and the probable variation in capacity over time. For the moment, only the problem of minimizing the cost of a given capacity (assuming that the compressors are operated 75 per cent of the time) will be considered without the complications introduced by the need for flexibility. The functions which have been written are therefore sufficient to determine the equilibrium or least-cost conditions after the costs of the several inputs have been specified.

### B. *Cost Functions*

The elements of cost in the long-run cost function may be divided into three categories: annual charges depending on the cost of installed equipment (interest, depreciation, property taxes, and obsolescence), annual operating costs depending on the quantities of the several capital goods installed, and other operating costs, which in this case are approximately proportional to the total cost. The nature of the production function under consideration is such that the only possibility of substitution among factors is between compressors and pipe. The quantities of labor, fuel, and maintenance are determined by the size and amount of the two types of capital equipment while raw materials (gas) are only consumed in the process as fuel. They may otherwise be left out of the calculations entirely since the input equals the output.

Cost data on the construction and operation of all interstate pipeline systems has been collected and analyzed by the Federal Power Commission.[9] Construction costs based on engineering experience are also available[1] for the period of the late twenties, when most of the pipeline systems now in existence were designed and built. These studies dictate the variables to be used in the cost function.[2]

varies. The present assumption of constant spacing (100 miles) simplifies the mathematical treatment and gives similar results as to substitution and total cost.

9. Federal Power Commission, *Natural Gas Company Cost Units*, 1945.

1. W. R. Kepler, "Gas Pipe Line Factors Affecting Minimum Cost," *American Gas Association Proceedings*, 1930, pp. 797–819 and H. C. Lehn, *op. cit.*

2. The statistical technique of the Power Commission study may be questioned because no correction was made for changes in construction costs and only simple correlations were used. However, the pronounced central tendency of all the cost distributions makes the errors due to this source of minor importance for the present purpose. All costs in both studies are stated in terms of variables derivable from equation (13): weight, diameter, and horsepower. The graphical results of the Power Commission study have been reduced to linear functions without introducing much inaccuracy over the range of observations.

These cost data may be summarized in the following cost function:

$$C = \pi (D, T, R, \overline{X}) = (a_1 i + b_1) H + (a_2 + a_4) iW +$$
$$(b_2 + a_5 i) D + a_3 + c_1 C \qquad (14)$$

Where $D$ is diameter, $T$ is thickness, $R$ is compression ratio,
  $H$ is horsepower (which depends on $X$ and $R$), $X$ is capacity,
  $W$ is weight of pipe (which is a function of $D$ and $T$).
  $C$ is the annual cost per 100 miles of line

$a_1 = a_1(H)$ = installation cost per horsepower
$a_2 = a_2(S)$ = cost per ton of pipe (as a function of working stress)
$a_3 =$        pipe installation cost depending only on length
$a_4 =$        pipe installation cost depending only on weight
$a_5 =$        pipe installation cost depending only on diameter
$b_1 = b_1(H)$ = annual operating cost per horsepower
$b_2 = b_2(T)$ = annual pipeline maintenance cost per inch of diameter
$c_1 =$        other operating costs as a fraction of total cost
$i =$          combined annual rate of interest, depreciation, and obsolescence

Although theoretically these parameters may be functions of the indicated engineering variables, the available data shows them to be practically constant over the range which has been experienced in practice. The following values of the parameters will be used as fairly representative of the pipelines built before the war. The chief inaccuracy is the use of 1945 operating data with prewar construction costs.

  $a_1 = \$130$ per Hp.
  $a_2 = \$100$ per ton
  $a_3 = \$300,000$ per 100 miles
  $a_4 = \$12$ per ton
  $a_5 = \$26,000$ per inch of diameter
  $b_1 = \$19$ per horsepower per year (compressors operated 6500 hours)
  $b_2 = \$1250$ per inch per year
  $c_1 = .18$

The determination of $i$ is a more difficult problem. If properly maintained, both pipelines and compressor stations will last indefinitely and only become obsolete if new operating techniques require higher pressures than the lines were designed to handle. The Federal Power Commission specifies depreciation rates of 2.5 per cent and 3.0 per cent for pipelines and compressor stations respectively, but these figures bear no necessary relation to physical life of the equipment if it is fully maintained. In the present study, the effect of

varying the combined rate of interest, depreciation, and obsolescence will be studied, assuming that interest and obsolescence are the dominant factors and that both pipe and compressors have the same combined rate, $i$.

## C. *Conditions of Equilibrium*

Given equations (13) and (14), the derivation of the long-run cost curve follows the textbook analysis with the engineering variables used in place of physical inputs for convenience. Hence only the difference between the two techniques require comment. The results of the solution of equations (7) using the data of the present example will be given first.

1. *The Least-Cost Solution.* Solution of equations (7) requires the partial differentiation of the production function and the cost function with respect to each of the engineering variables. The results of differentiating equations (13) and (14) are as follows:

$$\pi'_R = (a_1 i + b_1)\ \frac{\partial H}{\partial R}\ = 28.75\ (a_1 i + b_1)\ \overline{X} \tag{15}$$

$$\pi'_T = (a_2 + a_4)i\ \frac{\partial W}{\partial T}\ + 2820\ (a_2 + a_4)i\ \mathrm{D} \tag{16}$$

$$\pi'_D = (a_2 + a_4)i\ \frac{\partial W}{\partial D}\ + (b_2 + a_5 i)\ = 2820\ (a_2 + a_4)iT\ +$$
$$(b_2 + a_5 i) \tag{17}$$

$$\phi'_R = \frac{K_1\ D^{1.67}\ T}{R^2\ \sqrt{R^2 - 1}} \tag{18}$$

$$\phi'_T = K_1 D^{1.67}\ \sqrt{1 - 1/R_2} \tag{19}$$

$$\phi'_D = 1.67\ K_1 D^{.67}\ T\ \sqrt{1 - 1/R_2} \tag{20}$$

Any two of the equations (7) may be combined with the production function (13) to give the equilibrium values of the engineering variables. The solution is simplified by using the derivatives with respect to $T$ and $D$ first because it turns out that $T$ is a function only of the parameters:

$$\frac{\pi'_D}{\pi'_T} = \frac{\phi'_D}{\phi'_T}\ \text{from which}\ T = \frac{3}{2}\left[\frac{b_2 + a_5 i}{2820\ (a_2 + a_4)i}\right] \tag{21}$$

Assuming a value for $i$ of .06 (which is about the lower limit of the practical range of variation) and substituting the values of the other parameters, we can solve[3] for $T$:

$$T = \frac{1250 + 26,000}{2820\ (100 + 12)i} = .222\ \text{inches} \tag{22}$$

3. All calculations were made by slide rule.

We can take advantage of the fact that $T$ is a constant once the costs associated with pipe have been specified to rewrite the cost function in terms of the costs of one-dimensional physical inputs. This was not possible previously because pipe costs are a function of both diameter and thickness. If the latter is constant, they can all be expressed as a function of either diameter or weight. Equation (14) thus becomes:

$$C = HP_H + DP_D \text{ (omitting the constant elements } c_1 \text{ and } a_3)\quad(23)$$

where:

$P_H$ = annual cost per horsepower = $a_1 i + b_1$ = \$26.80

$P_D$ = annual cost per inch of diameter

$\quad = [(a_2 + a_4)\ 2820T + a_5]i + b_2$ = \$7030

The values of $R$ and $D$ can be most conveniently determined in terms of these two cost elements, $P_H$ and $P_D$.

Using the derivatives with respect to $R$ and $D$ and solving for $X$ gives:

$$X = \frac{K_2 D}{R (R^2 - 1)}\quad(24)$$

where

$$K_2 = \left(\frac{P_D}{48\ P_H}\right) = 5.47\quad(25)$$

This result can be checked by using the third pair of derivatives, which gives the same value of $K$. Eliminating $D$ between equations (24) and (13) gives the solution for $R$ in terms of the output, $\overline{X}$.

$$R^{2\cdot308} - R^{\cdot308} = \left[\frac{K_2^{\cdot77}}{(K_1 T)^{\cdot462}}\right] \overline{X}^{-\cdot308} = \frac{4.75}{\overline{X}^{\cdot308}}\quad(26)$$

$D$ is then determined from either (13) or (24).

The least-cost combination of inputs has now been specified for any value of $X$, since the secondary conditions for a minimum are satisfied by the derivatives. It is therefore theoretically possible to rewrite the cost function in terms of $X$ alone. Since this would result in a very unwieldy expression, a graphical solution is given in Fig. 1 below for the long-run cost curve. The other cost curves presented will be discussed later.

Since $T$ is a constant so long as the parameters are unchanged, we can express $D$ in terms of $R$ by eliminating $X$ between equations (13) and (24).

$$D = \left(\frac{K_2}{K_1 T}\right)^{3/2} \frac{1}{(R^2 - 1)^{2\cdot25}}\quad(27)$$

This result is analagous to the familiar "scale line."

2. *Effect of Price Changes on the Least-Cost Positions.* Since the general solution for the displacement of equilibrium due to a change in any cost element is quite cumbersome in the present case, only

FIGURE 1

COST CURVES: GAS TRANSMISSION
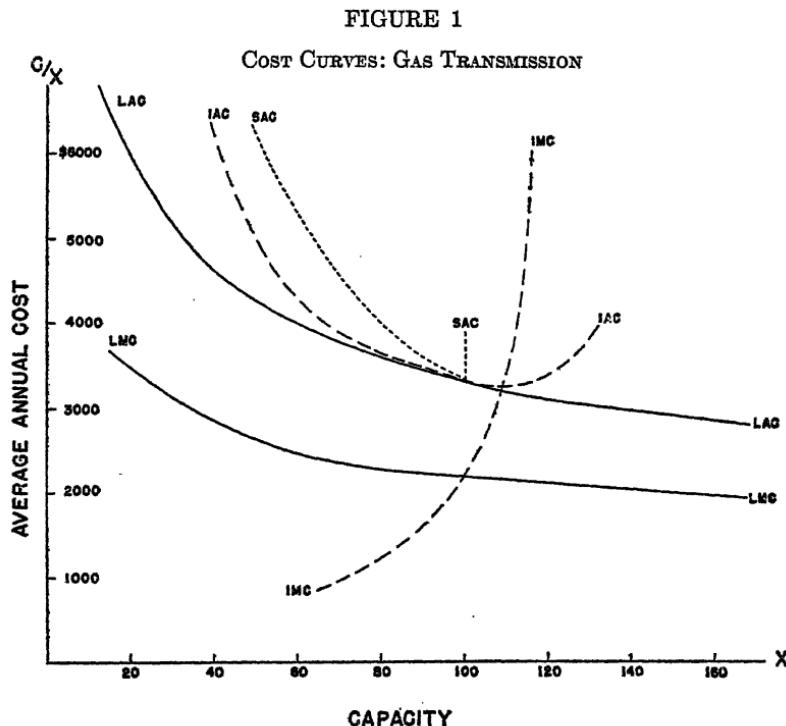


LMC: Long-run marginal cost
LAC: Long-run average cost
IMC: Intermediate marginal cost (pipeline fixed, horsepower varied)
IAC: Intermediate average cost
SAC. Short-run average cost

The long-run cost curves (*LAC* and *LMC*) were derived from equations (14) with $i$ equal to .06.

The intermediate cost curves assume that the pipeline determined by the long-run cost curve for $X = 100$ is fixed, and that $R$ is varied.

Short-run average cost was determined from the Federal Power Commission data assuming that the operation of compressor stations is the only cost which varies with output.

two examples of the effect of changes in factor costs will be considered. The variable elements are $i$ and the load factor.

In specifying the parameters used above, an annual rate of

operation of the compressors of 6500 hours or 75 per cent has been assumed, which represents a high load factor. If instead the demand for gas is so variable that an annual rate of operation of only 25 per cent can be expected, the annual cost per horsepower is reduced about 40 per cent. Since the costs of maintaining the pipeline are not reduced at all by the lower load factor, the least-cost solution will require more compressors and less pipe.

A second variable which affects the least-cost position considerably is the combined rate of interest, depreciation, and obsolescence, $i$. The expected life of the gas field, demand for gas, the rate of interest, and even the property tax will all affect the value of $i$ to be used. We shall consider the effect of a rise in $i$ from 6 per cent to 10 per cent. This change will have the same effect on the optimum $P_D/P_H$ ratio as the fall in the cost of horsepower assumed above because pipe is considerably more capital-intensive than compressors.

The optimum ratio of horsepower per unit of output (which is proportional to $R$) to diameter has been shown to depend only on the $P_D/P_H$ ratio and the thickness, $T$. As a final case let us consider the effect of doubling the ratio $P_D/P_H$, which might be caused by number of relative price changes among the parameters. If the interest rate is not among the elements which change, we can assume $T$ to remain constant (which is of assistance in graphical representation).

These three assumptions, together with the original data (Case 1) may be summed up in the following table:

|        | $i$ | $T$ | $P_D/P_H$ | $W/H$ (for $X = 100$) |
|--------|------|-------|------|------------------|
| Case 1 | .06 | .222 | 2.62 | 6.14 tons per hp. |
| Case 2 | .06 | .222 | 3.64 | 5.41 tons per hp. |
| Case 3 | .10 | .183 | 3.00 | 5.24 tons per hp. |
| Case 4 | .06 | .222 | 5.24 | 4.74 tons per hp. |

The magnitude of the resulting shifts in the optimum combination of factors is shown approximately by the ratio of weight of pipe to horsepower. A more complete presentation of the results is made in the graphical solutions below.

3. *Graphical Solutions.* Engineering production functions permit graphical solutions in terms of each of the three sets of variables. Each has advantages for particular uses, and together they shed light on the calculations underlying the production function in terms of capital and labor which the economist is accustomed to use. The present example is a favorable one for graphical analysis because so

long as the rate $i$ does not change, the engineering variables may be reduced to two, diameter and compression ratio.

The production function in terms of the engineering variables (Fig. 2) is likely to be the simplest mathematical relationship, as it

FIGURE 2

ENGINEERING PRODUCTION FUNCTION: GAS TRANSMISSION
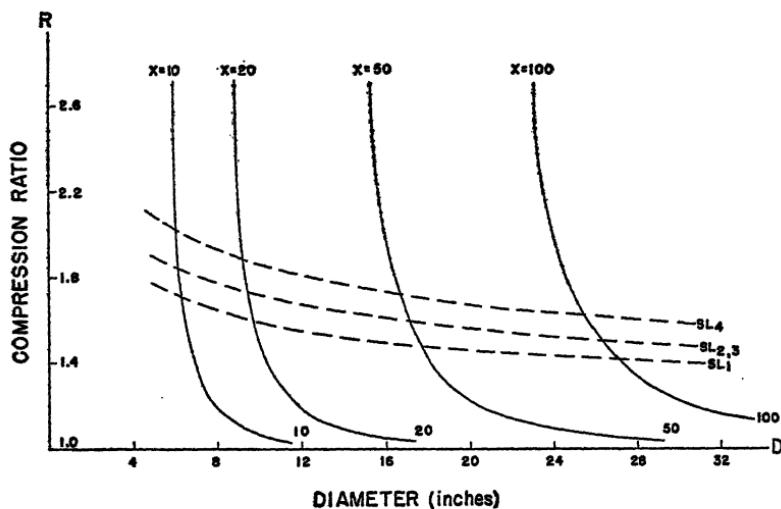


is in the present case. However, since the isocost lines are not straight in terms of these variables, a transformation to dimensions which have constant costs is desirable. Fig. 3 shows the production function in terms of horsepower (for which a constant unit cost has been assumed) and weight of pipe for a constant thickness. While a change in thickness will only require a renumbering of the isoquants in Fig. 2, it will change the values of $H$ and $W$ and hence require a separate cross-section in the $H$-$W$ plane for each value of $T$. Fig. 3 can therefore be used to compare only price changes involving no change in $i$, since the optimum thickness varies with $i$.

The relation between generalized factors of capital, labor, materials, etc. with which the economist usually starts is in general the most difficult to derive. For this reason it comes at the end of the present analysis rather than at the beginning. The reason of course is that the function relating capital, labor, etc. to ouptut is defined implicitly by the functions which we have analyzed but cannot be written explicitly in the present case. When the two preceding functions have already been derived, the only need for

transforming the input combinations into more general categories of
capital and labor is to analyze the demand for these general factors.
Since capital alone represents about two-thirds of the total cost in
the present process, it has been treated as a separate factor and all

FIGURE 3

PRODUCTION FUNCTION (PHYSICAL INPUTS)
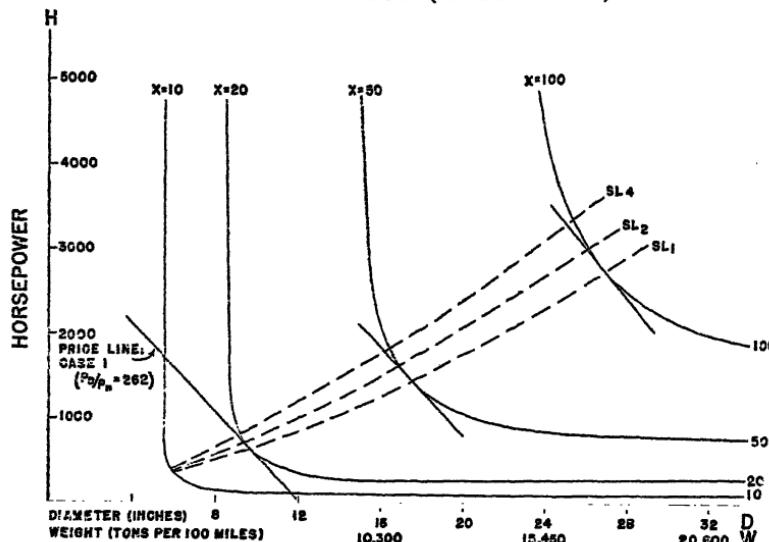


current inputs (including labor) lumped together as materials. This
procedure is valid only if the relative prices of current inputs do not
change. With this assumption value units can be used to measure
the quantities of both capital and materials. Their price ratio is of
course $i$. The results of such a transformation are shown in Fig. 4
below. The increase in $i$ assumed in Case 3 is shown to cause very
little decrease in the amount of capital used because in this neighbor-
hood the elasticity of substitution of capital for other inputs is very
low.

Scale lines are drawn on all three production functions to
represent the least cost positions for various values of the cost
parameters. Those in Fig. 3 can be determined graphically for any
assumed value of the price ratio of horsepower to weight of pipe.
Those in Fig. 2, however, cannot be determined graphically since
prices are not uniquely related to values of the engineering variables.
On this diagram, therefore, the scale line is determined from equation
(27) and represents values of the engineering variables at the least-
cost solution.

The method of solution was indicated in section II above. Using the Lagrange multiplier with two constraints, output and materials constant, gives three equations of the following form:

$$K_i' - \lambda_1 M_i' - \lambda_2 X_i' = 0 \qquad (i = D, T, R)$$

where $K = a_5 D + (a_2 + a_4) W + a_1 H$ = the quantity of capital

$M = b_2 D + b_1 H$ = all current inputs (materials)

FIGURE 4

ECONOMIC PRODUCTION FUNCTION
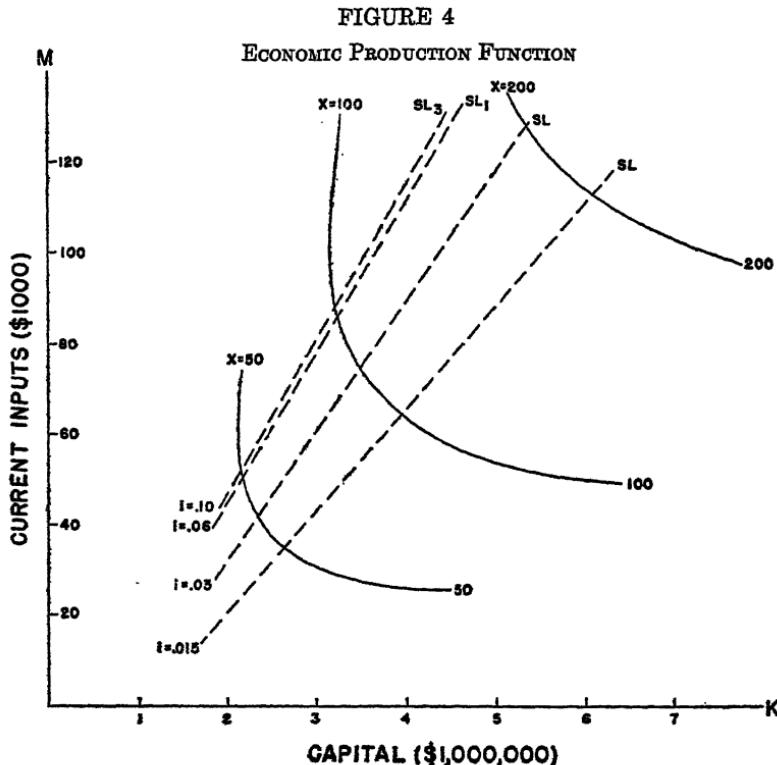


Eliminating the two undetermined parameters $\lambda_1$ and $\lambda_2$, then eliminating $T$ by means of the production function gives:

$$17,450 \, D^{2/3} \, \sqrt{R^2 - 1} = \frac{80,200 \, \overline{X} \, R}{D} - \frac{276,000}{(R^2 - 1)}$$

This can be solved for particular values of the two variables $X$ and $M$ in conjunction with the definition of the materials factor:

$$D + .436 \, R\overline{X} = \frac{\overline{M}}{1250} + .211\overline{X}$$

The values of $D$, $R$, and $T$ from this solution are substituted into the equation defining the quantity of capital to give the isoquants in Fig. 4. The scale line was determined mathematically for $i = .06$ and graphically for the other values.

FIGURE 5

THE PRODUCTIVITY OF CAPITAL



CAPITAL ($1,000,000)
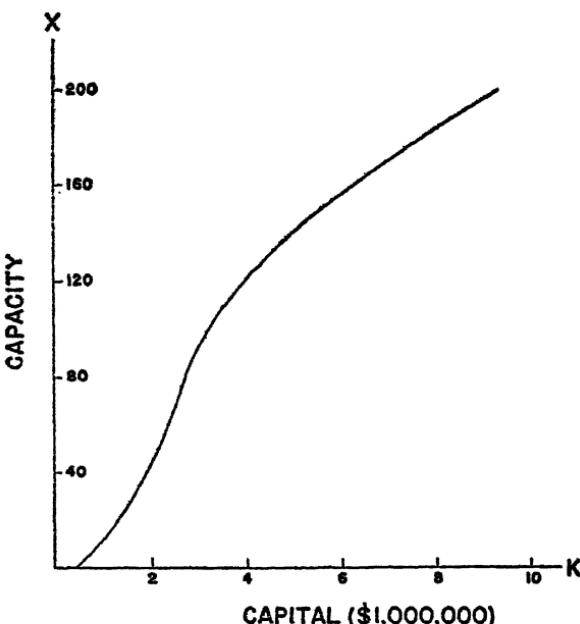
Since treatment of all current inputs as a single factor, materials, is valid only so long as their relative prices do not change, the quantity of materials used here depends on the particular prices assumed.

The productivity of capital, Fig. 5, is the particular solution of the above set of equations given by holding $M$ constant at $90,000 and varying $X$.

IV. THE USE OF ENGINEERING PRODUCTION FUNCTIONS

A. *Results of the Present Study*

The essential and perhaps rather unusual feature of the transmission of gas is that the two principal elements of the technique — pipe and pumping engines — can be substituted for each other over a considerable range. From Fig. 3 it can be calculated that a deviation from the optimum ratio of horsepower to weight of pipe of 50 per cent will cause a rise in cost of only 10 per cent. This char-

acteristic of the technique gives to natural gas transmission a considerable degree of flexibility. A pipeline which is ultimately intended to provide a capacity of 100 million cubic feet a day may initially have only enough compressors installed to pump 60 million, but the average cost will be less than 10 per cent greater than for the most efficient line which could be built to provide a capacity of 60 million (see Fig. 1). The very marked economies of scale of large pipelines are equally responsible for this flexibility, which of course is only in one direction. Once the optimum compression for 100 million capacity has been installed, the only variable costs, if output drops off, are fuel and a small part of maintenance. The plant cost curve rises very rapidly and at 60 million is 40 per cent higher than that of the most efficient installation.

These three cost curves, two of which result from the present type of production-function analysis, would of course be invaluable for economic analysis if they were more generally available. Of the three, the intermediate curve, representing the possibility of varying the more flexible capital factor, is for some purposes more useful than the plant curve. If the pricing of the services of an existing pipeline were controlled by marginal cost, for example, this would be the relevant curve. In the present case, the marginal cost rises very rapidly beyond the minimum-cost point, and marginal-cost pricing might result in a slow rate of growth and high prices. Another important use of this production function is in the analysis of investment. The cost curves suggest that if demand is expanding, considerable economies will be effected by building ahead of demand.

Turning from the industry to the economy as a whole, we are interested in the demand for factors of production and the effect of changes in their relative prices. The present analysis has isolated only the most important factor, capital. The scale line in Fig. 4 shows that when capital charges are at a practical minimum — 6 per cent for interest, depreciation, and obsolescence — the cost of capital is more than twice the total of all current operating costs ($M$); this ratio decreases slightly with increasing size. Although the two principal inputs, pipe and compressors, have an elasticity of substitution of about .37 in the range considered, they are both quite capital-intensive. Therefore the elasticity of substitution of materials, fuel, and labor for capital is very low for values of $i$ between 5 and 12 per cent or so, which is the practical range in the United States.[4] If obsolescence and risk were eliminated, however, there

4. The minimum value of $i$ suggested by one gas company executive at the present time is 10 per cent, consisting of the following components: depreciation, 2.5 per cent; property taxes, 1.5 per cent; return on investment, 6.

would be a considerable possibility of substituting capital for current inputs.

Since the shape of the total productivity curve has been a matter of academic debate, the product curve for capital which is derivable from the isoquants of Fig. 4 is presented in Fig. 5, for the range over which the postulated production function is not too inaccurate. It would curve downward if extended, and in general it shows the familiar textbook shape, as do the isoquants from which it is derived. The possibility of substituting one factor for another in industrial processes typically involves such a continuous change in the form of the inputs as it does here. The fixed production coefficients which have been postulated here do not, of course, produce any discontinuities in the isoquants so long as some substitution between two capital factors such as pipe and pumps is possible.

Lest the data given here be used unaltered for a present-day discussion of the gas industry, it is desirable to mention the matter of technological change. Two important changes have revised the parameters in the production function of the thirties. Allowable working stresses in pipe have doubled due to better pipe-making techniques and more experience with large diameters. This merely changes the constant $K$ used above, which affects the shape of the function in terms of capital and other factors. An equally important development has been the modification of Weymouth's formula at high pressures due to the deviation of gas behavior from that predicted by Boyle's law. The increase in quantity for a given pressure may be as much as 20 per cent for pressures above 1000 pounds. The effect on the production function of these two changes together, is to shift the equilibrium solution from the minimum thickness with large diameter (i.e. low pressure) to very high pressures (and smaller diameters) for values of $X$ less than 200 million feet per day.[5] Hence many existing low-pressure lines may become obsolete. Since the phenomenon of supercompressibility introduces a factor into the production function which greatly complicates it mathematically, the present paper has been limited to the low-pressure transmission techniques which have been standard until recently. Despite these changes, the generalizations about flexibility and economies of scale which have been made are unaffected.

5. See "A Study of the Economics of Gas Pipe Line Design," *Proceedings, Natural Gas Section, American Gas Assn.*, 1942, p. 201 where the optimum values of the different variables under modern conditions are derived. For quantities in excess of 250 million feet per day, the optimum values are the same as those derived from Weymouth's formula here — minimum wall thickness and operating pressure, resulting in large diameter.

B. *Further Applications*

Now that a technique for utilizing engineering results has been demonstrated, several suggestions can be made for its application to other industries. First it is necessary to find processes or units of equipment which are sufficiently important to require detailed analysis. In the case just studied there were two such units, pipe and compressors, which accounted for about 75 per cent of total cost. Since the production function of the compressor contained only one engineering variable, it was possible to consider it as merely a way of transforming compression ratio into horsepower and hence as determining the cost of a given compression ratio. The pipeline production function in its most general form contained five variables, which were reduced to three by making simplifying assumptions.

Production functions have been derived by similar methods for a number of processes involving different types of technology, including evaporation, electrolysis, electric power transmission, and comminution. In no case was it necessary to consider more than three dimensions of the process as variables in a realistic application. These processes have in common the fact that they can be described satisfactorily by an analytical model based on the laws of electricity, thermodynamics, chemistry, or mechanics. The production functions which can be derived for them are therefore of a very general nature and can be made applicable to a variety of industrial plants by substituting the proper values of the parameters. Processes are building blocks which can be combined in particular plants in many possible ways. Hence a more general knowledge of production is secured by analyzing separate processes than by studying individual plants, which represent only one possible combination of them. The number of types of plants is much greater than the number of processes.[6]

This method of analysis solves some problems but raises new ones. The first is the analysis of cost data in terms of the engineering variables. The present example was chosen because such cost data was available. Equipment costs are normally stated in terms of various dimensions which occur in the production function, but it is more difficult to ascertain the variation in maintenance and operating costs on the same basis. Since engineering analysis suggests the variables to be used and often the form of the equation to be fitted,

6. See R. N. Shreve, *The Chemical Process Industries*, Chap. II for a discussion of the use of unit processes in analyzing plants in the chemical fields and also R. Schuhmann, "The Unit Processes of Chemical Metallurgy," *American Institute of Mining and Metallurgical Engineers, Technical Publication* No. 2363, June 1948.

however, statistical determination of the cost function is simplified.

The economist, who is accustomed to think in terms of whole plants, may be skeptical of the value of a production function for only part of one. In gas transmission little violence was done to the facts by concentrating on only two processes because the other cost elements are pretty well fixed by the volume of gas transported and are in any case unimportant in determining substitution and economies of scale. In more complicated plants, however, it may be necessary to consider five or ten processes to get a realistic picture of productive possibilities. Fortunately, the necessity of combining all these processes in a co-ordinated unit reduces the final number of variables. Many of the engineering variables will be common to several processes; if a limitation occurs in one process it must be followed in all. Once the engineering variables which affect several processes have been fixed, each process can be treated as a separate unit and its own variables determined independently. Concentration on a particular aspect of the production function, such as substitution or scale effects, will permit a further reduction in the number of variables. At worst, the economist can limit himself to the two or three principal processes involved and consider the others as fixed by them. This is a common method of engineering approximation used in planning.

Finally, we come to the meaning of the production function itself. In only a relatively small proportion of cases do engineers understand well enough what goes on in the transformation of materials into products to reduce the transformation to an analytical model. Yet without such a model, the only productive possibilities which are known precisely are those which have been tried on a commercial scale. It is possible to produce an isoquant for any size of pipeline on analytical grounds, but the only combinations which have any accurate meaning for blast furnaces or turbines are those which have actually been built. In a sense technological change is involved whenever a new motor is designed, even though no new engineering principles are involved, merely because the outcome is uncertain. Therefore we must attach increased degrees of technological risk to points on the production function which deviate from the range which has been covered by actual plants. It is this variation in technological risk which makes the movement along some production functions in response to price changes much slower than the movement along others, Each new point involves some degree of developmental expense, and the organization of the industry determines to a large degree the willingness of entrepreneurs to undertake this risk.

## C. *Conclusions*

The contrast between the present method and existing statistical methods should be clear. They are essentially complementary. The few statistical attempts which have been made to determine production functions pertain only to particular plants and short-run variations.[7] Where no complete engineering explanation of the process exists, we can only hope to get such short-run functions with the variables and perhaps the shape of the equations indicated by engineering considerations. Where technology is further advanced,[8] as in chemical, aeronautical, and electrical engineering, we can derive long-run production functions of a more general nature. It is only in such cases that there can exist in the entrepreneur's mind more than a hazy idea of untried variations in productive combinations which may be feasible.

Even if we are interested only in the productive possibilities with unchanged prices, the cost curve derived from a production function by the present method has distinct advantages over the statistical variety. In the usual method the relevant variables, the shape of the equations, and the values of the parameters must all be determined from the same set of observations, whereas in the present case, only the parameters need to be determined. Therefore rather crude cost data can be used without great error in the shape of the cost curve. It is only with regard to the nontechnological elements in cost that the results depend entirely on the statistical observations.

<div align="right">HOLLIS B. CHENERY.</div>

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

7. Cf. W. H. Nicholls, *Labor Productivity Functions in Meat Packing*, and Anne P. Grosse, *The Technical Production Function and Cost Minimization in Basic Open Hearth Steel Making*, Radcliffe Ph.D. thesis, 1949. The latter uses engineering information to determine the variables to be considered and is based on experimentation with existing furnaces as well as commercial operation.

8. Samples of engineering literature which deal with analytical production functions are the following:

W. H. McAdams, *Heat Transmission*, Chap. VII.

Seelye, *Electrical Distribution Engineering*, Part III.

R. Von Mises, *Theory of Flight*, Part IV.

C. Tyler, *Chemical Engineering Economics*, Chap. VII.

Hetherington and Huntington, "Economics of Gasoline Plant Design," *Refiner and Natural Gasoline Manufacturer*, August 1941. This gives a good example of the determination of the optimum value of an engineering variable which affects several processes.

# THE PHYSIOCRATS' CONCEPT OF ECONOMICS

## SUMMARY

Introduction, 532. — I. The nature and scope of physiocracy: a normative and mathematical science based on "natural law," 534. — II. The theory of innate economic ideas, 538. — III. A mathematically exact science deduced from first principles, 540. — IV. Quesnay's own view of the new science, 544. — V. Quesnay's use of the inductive method and of mathematics, 549. — VI. The gulf between Quesnay and his disciples, 551.

Historians of economic thought generally hold that the Physiocrats were founders of "the first strictly scientific system of economics."[1] There was economic thought before the Physiocrats, to be sure, but this previous thought consisted of scattered theories, like Gresham's law, and it was generally treated as part of ethics or of politics. Such subordination of economic thought to ethics is only to have been expected in a medieval and early modern society which looked upon social activity as moral and therefore subject to the ethical teaching of the Christian churches. It was likewise natural that economic thought be considered a branch of politics in the sixteenth and seventeenth centuries when the absolute state had established itself in European society and succeeded in embracing or controlling most facets of human activity.

In the mid-eighteenth century, however, in that period known as the "Age of Enlightenment," there occurred an intellectual revolution which was directed against the traditional control of both Church and State over social activity. And out of that revolution there seemed to emerge a new science. Such, at least, was the pronouncement of an anonymous Physiocrat who wrote that "economic science has at last penetrated the sanctuary of the Muses . . . First because of the importance of its object, in the chronological order it is the last human branch of knowledge to be discovered."[2] This seems to have been the general opinion of men of the Enlightenment. The gossipy Baron Grimm, for example, wrote from Paris to his German subscribers that "political economy has become *la science à la mode* in

1. The phrase is Auguste Oncken's. See his compilation, *Oeuvres économiques et philosophiques de F. Quesnay* (Paris, 1888), ix. It is a statement that is as universally agreed to as any broad generalization in the history of economic theory.
2. "De l'utilité des discussions économiques," in *Physiocratie* (Paris, 1768), IV, 3. *Physiocratie* is a six volume collection of essential physiocratic writings made by Pierre-Samuel Dupont. Authorship of the individual articles is not indicated in the collection.

France," and he complained that "a sect [the Physiocrats] wishes to dominate this science."[3]

Certainly the Physiocrats believed that they had created a new science.[4] They thought not only that their leader, François Quesnay, had discovered it but also that he had developed all the essential truths it could properly contain. Mirabeau considered Quesnay's *Tableau économique* one of the world's three great discoveries — equalled only by the invention of printing and the discovery of money.[5] The Physiocrats were all content to sit at the feet of their master who, like Confucius,[6] was the oracle who possessed the truth in all its fullness. For them physiocracy was a new discovery which, like Minerva springing from the head of Zeus, was born full-blown, completely developed. The Physiocrats thought, therefore, that their appointed task was to explain and popularize the truths discovered by Quesnay. None of the followers thought that they, themselves, had anything original or personal to contribute to the development of economic theory. The first ensemble of physiocratic doctrine, Mirabeau's *Philosophie rurale*, was built around Quesnay's *Tableau économique* with the express purpose of popularizing it.[7] Mirabeau wrote at Quesnay's suggestion, to some extent under his supervision, and he included at least one chapter written by Quesnay himself. In like manner, Le Mercier's *L'Ordre naturel et essential des sociétés politiques*, which turned out to be the classic presentation of physiocratic theory, was written in Quesnay's apartment, under his direction, and with the purpose of presenting the doctor's theories in systematic, orderly fashion.[8]

3. *Correspondance litteraire, philosophique et critique* (Paris, 1813), Pt. I, Vol. V (February 1766), 480.

4. The best source of information on the Physiocrats and their contemporaries is Dupont's "Notice abrégée des différents écrits modernes qui ont concouru en France à former la science de l'économie politique," published in eight numbers of the physiocratic journal, *Éphémérides du citoyen*, in 1769.

5. Cited by Louis de Lomenie, *Les Mirabeau: Nouvelles études sur la société française au XVIIIe siècle* (Paris, 1879), II, 311.

6. Quesnay was likened to Confucius, Socrates, Zeno and other such intellectual giants whose teaching was enthusiastically received by their disciples, who in turn became missioners of the great man's gospel.

7. Mirabeau's book, published in 1763, was a miserable failure. It did not offer the age the brilliant phraseology and the cystral-clear arrangement of ideas then in vogue. It was because of Mirabeau's failure that Quesnay looked to Le Mercier for a successful popularization of his ideas.

8. Contemporaries, both the Physiocrats and their opponents, all looked upon Le Mercier's book as the definitive statement of physiocratic doctrine. Dupont called it a "sublime book" in which the "truths discovered by Dr. Quesnay are so superiorly and so clearly developed." (*Physiocratie*, III, 15.) Adam Smith believed it was the "most distinct and best connected account of this [physiocratic] doctrine." (*Wealth of Nations*, Bk. IV, Chapter IX.)

So the Physiocrats considered themselves a "school" of theorists who differed from each other on no important doctrine. This was an age when "sects" and "systems" had fallen into intellectual disrepute, and the Physiocrats had to face the charge of blindly following a man "who had thought only of the operations of surgery and the rules of medicine, to which he added for his amusement a confused metaphysics,"[9]   But they were willing to face the charge because they were convinced of the correctness of Quesnay's doctrine — and because they thought all intelligent persons would some day come to agree with them in accepting it.   Dupont tried to parry the accusation that the Physiocrats were a "sect" with countercharges which evaded the point at issue.

And now a word to the severe enemies of *sects*. If those who regard all men as their brothers, who occupy themselves peacefully and incessantly with developing their interests, their duties and their rights, who show that there are sacred and supreme natural laws, the notion of which is evident to all who reflect on it and the sanction of which is visible, pressing, imperious and inevitable, . . . if such a group is called a sect who prove methodically all these things by calculation and by measure, then we well deserve to be hated, decried and persecuted by the wicked, by violaters of the rights of others, by breakers of the natural law, by arbitrary despots and tyrants.[1]

Le Trosne handled the question better by admitting frankly that the Physiocrats "form a sect, if it suffices for this [designation] to have the same opinions, the same language, to be perfectly in accord on all points, to recognize a single master."[2]   Both the Physiocrats and their opponents, then, considered the group a small school of men sitting at the feet of a single master, men whose function was to popularize the master's thought and thus bring the blessing of the new science to the entire world, a group, therefore, in which there could be no disagreement on important points of doctrine, for Quesnay's writings were the "inspired word" to which they all looked for ultimate truth.

## I

What did the Physiocrats think of this wonderful new science? What was its object?   Its nature?   Its method?   What was it to embrace?   How was it to discover all facets of social truth?   What,

9. Abbé Gabriel Bonnet de Mably, "Du commerce des grains," *Collection complète des oeuvres de l'abbé de Mably* (Paris, 1794–95), XIII, 295.

1. From the *Éphémérides du citoyen* of 1769, the official physiocratic journal edited by Dupont.   The selection is found in Eugène Daire (ed.), *Physiocrates* (Paris, 1846), 315–316.   Daire's collection of physiocratic writings remains the most complete in any language.

2. *De l'ordre social* (Paris, 1777), 312.

if anything, was to be put aside as foreign to itself and properly the object of another science?  These are questions the Physiocrats had to answer for themselves in an age of rationalism so that their new science could obtain a recognized and clearly defined position.  The answers they formulated are of some importance in the history of economic theory, for it is partly upon their answers to these questions that the Physiocrats are to be considered either the first contemporary economists or the last of the earlier modern economists.

Writing, as they did, in an age of intellectual confusion,[3] the Physiocrats did not have as clear-cut a picture of the scope and the object of economics as did their successors of the early nineteenth century.  In the first place, they did not cut entirely free from the older concepts which put economics down as a branch of ethics or of politics.  They called themselves *philosophes économistes*, and they looked upon their new science as a social philosophy including economic, political, ethical, and social activity.  It was a normative social science, apparently, for it was based on natural law, as the word *physiocratie* indicates.  Baudeau significantly entitled the work in which he tried to sum up physiocracy *Introduction à la philosophie économique*, and he concluded this work by stating, "*Voilà le droit naturel et la philosophie morale*," a science which is to be perfected by "*l'instruction morale économique*, that is to say, instruction in the natural law of justice in its essence."[4]

The Physiocrats bound their new science tightly to "natural law," sometimes making it a part of the laws of nature,[5] sometimes identifying the two.  Dupont states, for example, that "with a little reflection one can see with certitude that the sovereign laws of nature include the essential principles of the economic order."[6]  Again, he writes to J.-B. Say that economics "is the science of natural law, applied, as

3. It is important that one remember how confusing things intellectual were in this age when almost every thinker tried, above all, to bring order to his subject.  It is an age when Cartesian rationalism, long accepted by advanced thinkers, has fallen into decline.  Cartesianism becomes the officially accepted philosophy of the Sorbonne, however, only in this age, replacing the even more decadent scholasticism.  Empiricism, imported from England, had come to be accepted by the more daring thinkers by 1775.  Romanticism, meanwhile, was beginning to undermine any form of rationalism.  It should also be remembered that these methods and these systems were not mutually exclusive in all respects.  It is difficult, therefore, to find any one thinker who uses one system and belongs to one school exclusively.

4. Daire, *op. cit.*, 819–820.

5. The younger Physiocrats are not consistent in differentiating between "natural law" as a moral law and "law of nature" as a Newtonian physical law.  The two concepts have merged in their minds, as we shall indicate later in this article, and it is only occasionally that they distinguish them from each other.

6. In Oncken, *op. cit.*, 362.

it should be, to civilized society."[7]   This same definition is offered by the author of the article *De l'utilité des discussions économiques* when he states that economics is "nothing but the application of the natural order to the government of society."[8]

Although Quesnay makes no specific statements on the nature or the scope of economics, the younger Physiocrats thought their definitions were implied in his writings and were derived from his conversations.[9]   In his *Despotisme de la Chine*, which was the closest Quesnay ever came to arranging his physiocratic teachings into a unified piece, he tells how in that ideal country of China "ethics and politics form a single science."[1]   This science is the model for his own physiocracy.   Quesnay goes on to tell how the Chinese U-King, apparently a combination constitution and bible, governs all man's social and religious conduct.

> These sacred books [of U-King] include a complete ensemble of religion and the government of the empire, of civil and political laws; both are dictated irrevocably by the natural law, the study of which is very searching and is, indeed, the capital object of the sovereign and the scholars charged with the details of administration.[2]

So the new science, which is based on natural law, is wide in scope.   Apparently it is to include all social conduct, all of man's dealing with man.   This was held as late as 1815 when Dupont condemned Say for restricting economics to the "science of wealth, which is only a collection of calculations."[3]   "You see, my dear Say," Dupont went on, "that our science is very extensive, that it includes a great number of objects.   Why do you restrain yourself to that of wealth? . . . Your genius is vast.   Do not imprison it in the ideas and the language of the English. . . . Economics is the science of enlightened justice in all its domestic and foreign social relations."[4]   In 1777, when the physiocratic school was twenty years old and there was little doubt as to its official teaching, Le Trosne had indicated the scope of physiocracy in more specific terms: "This science, taken

7. In Daire, *op. cit.*, 397.
8. *Physiocratie*, IV, 9.
9. Gustave Schelle, who has done more work on the activity of the physiocratic school than anyone else, observes: "It was in his [Quesnay's] apartment that economics was founded, more by Quesnay's conversations than by his writings." *Le Docteur Quesnay* (Paris, 1907), 122.
1. In Oncken, *op. cit.*, 605.   Quesnay's *Despotisme de la Chine* has recently been translated into English and is published as the second volume of Louis A. Maverick's *China, A Model for Europe* (San Antonio, 1946).
2. *Ibid.*, 605.
3. In Daire, *op. cit.*, 397.
4. *Ibid.*, 397, 415.

in its ensemble, includes all parts of the administration: agriculture, commerce, industry, taxation, justice, the police, legislation, peace and war: everything that is related to the security, the tranquillity and the welfare of men is its province."[5]

The Physiocrats, then, thought that they had discovered a new science, that it was an elucidation of natural law, and that its scope extended to all of man's dealing with man and with nature. It was therefore a moral science governing man's social activity, much the sort of thing that John Locke once hoped to achieve for ethics by applying to that subject the laws discovered by his friend Newton. But physiocracy was also thought to be an exact science. Just as God had reserved to the French (in the person of Descartes), the honor of making an exact science out of philosophy, Dupont told the readers of *Éphémérides du citoyen*, so the honor of molding morality and politics into a mathematical science had been given to the same people in the person of Quesnay.[6] That he was stating the official physiocratic position is evident from all their references to the new science, "a true science," says Baudeau, "which perhaps does not cede anything to geometry itself."[7] "Economic science," according to another Physiocrat, "is as constant in its principles and as susceptible of demonstration as the most certain physical sciences."[8] And Le Trosne calls it "an exact science subject to invariable rules."[9]

In thus defining and describing physiocracy the Physiocrats appear to have put themselves in an anomalous position. They can consistently believe it to be both normative and mathematical only if they hold that human conduct is susceptible to the same physical laws that govern the physical universe. Such a deterministic belief was not unknown in the eighteenth century, and although Quesnay had specifically written in defense of freedom of the will[1] and had written nothing directly to encourage a deterministic approach to economics, nevertheless the younger Physiocrats all show a tendency to eliminate the variable human factor from their calculations. They were convinced that Quesnay had discovered the secret that would make all men act rationally in the future: the utility of conforming

5. *De l'ordre social*, 346–347.
6. In Oncken, *op. cit.*, 716.
7. *Première introduction à la philosophie économique*, in Daire, *op. cit.*, 655.
8. "De l'utilité des discussions économiques," *Physiocratie*, IV, 9.
9. *De l'ordre social*, 320.
1. His most extensive defense of free will is his chapter "La liberté" in his *Essai physique sur l'économie animale*. His stand here is essentially the same as that of the medievalist schoolman. He also asserts freedom of the will in his *Droit naturel* and *De l'immortalité de l'âme*. These are all in Oncken's collection of Quesnay's writings.

to "natural law." This, said Dupont, was Quesnay's great contribution to humanity;[2] "a principle of the greatest fertility," Le Trosne called it, "which decides all questions of political economy, dissipates all prejudices, undergoes neither exception nor modification."[3] The philosophical problem of free will was therefore unimportant to the Physiocrats,[4] for they believed they had discovered the law of self-interest to which all men were as much subject as a stone is to the law of gravity. They were therefore unembarrassed by proposing that their science which dealt with human actions was both normative and mathematical.

## II

Each science imposes its own proper discipline upon those who work in it. Each science has its own method of reasoning, its own rules of procedure, its accepted way of arriving at valid conclusions. Perhaps the Physiocrats' concept of economics can be clarified, then, by seeing what method of reasoning they considered proper for the elucidation of economic truths. As typical Cartesian rationalists,[5] the Physiocrats subscribed to the theory of innate ideas. "Our knowledge of this law [natural law, which includes all economic laws]," Le Mercier writes, "is written in all our hearts."[6]

> The justice and the necessity of these natural laws are of a certitude which they themselves show to all men, without the help of any sensible sign. . . . It is in the code of nature itself that they are found written, and we distinctly read them all there with the aid of reason, this light which *illuminat omnem hominem venientem in hunc mundum.*[7]

Dupont goes even further and insists that economic truths manifest themselves to the simple savage as fully and certainly as to

2. Dupont states this in a letter to J.-B. Say. See J.-B. Say, *Cours complêt d'économie politique* [17th. ed., edited by Horace Say (Brussels, 1844) ], 582.

3. *De l'intérêt social,* 713. This second work of Le Trosne's is published in the same volume as his *De l'ordre social* (Paris, 1777).

4. The younger Physiocrats did not deny free will. They dismiss it quickly, however, to concentrate on what they call "social" or "physical" freedom, which is the liberty to do as one likes with his property.

5. It is not generally appreciated that Cartesianism was the "new" and "daring" philosophy in France until the middle of the eighteenth century. It did not receive official status at the University of Paris until after the de Prades scandal of 1751–52. Only the most "advanced" thinkers in the 1760's had revolted from the Cartesian system in favor of English empiricism, presented most effectively to the French thinker by Condillac's *Traité des sensations* (1754) and Helvetius' *De l'esprit* (1759).

When we say the Physiocrats were Cartesian rationalists, exception must be made for Quesnay, and to some extent for Mirabeau, as we shall show later in this paper.

6. *L'ordre naturel et essential des sociétés politiques* (London, 1767), II, 434.

7. *Ibid.,* I, 120–121.

the educated man. "These evident principles of the most perfect constitution of society," he says, "manifest themselves, of themselves, to man. I do not mean to say only to an instructed and studious man, but even to the simple savage."[8] In the same vein the physiocratic journal, *Éphémérides du citoyen*, condemned the Italian economist and criminologist Beccaria for using the inductive method in the sociological sciences.

We can know these sciences [the moral, political, and economic sciences comprising physiocracy] in their full extent, because their fundamental principles are by nature quite evident to those of us who wish to reflect a bit, and sometimes even despite ourselves. In applying ourselves to a thorough knowledge of these principles and always taking them as our point of departure, we arrive easily and with the greatest certitude at their most remote conclusions: an invincibly clear logic conducts us there rapidly by a series of incontestable deductions.[9]

Le Trosne, writing in 1777 when physiocracy was fully formulated as a science, handles the question of method and the origin of knowledge of economic principles more fully than does any other Physiocrat. In the introductory chapter of *De l'ordre social* he likens knowledge of economics to a "light" which floods the soul of the economist. "Struck by the certitude of these principles," Le Trosne continues, "he is firmly convinced that they will one day be made manifest to all men."[1] Economic laws are known innately because economics is "a science derived from the first principles of justice which enlighten all men."[2] It is not made known to man by a particular act of revelation on God's part; instead "it is simple, certain, laid bare to all eyes, it is written in obvious characters in the great book of nature."[3]

The Physiocrats seem confused in referring now to "the great book of nature" and then to "the heart" as the repository of economic truth. One naturally asks whether the Physiocrats looked for truth outside themselves in "nature" or in their own minds. The solution of this difficulty lies in the age's identification of "nature" with one's own mind. For it is by reflection rather than by observation that one reads in the "great book of nature."[4] Le Trosne demonstrates

8. "Discours de l'éditeur," *Physiocratie*, I, xix.
9. 1769, VI, 62. Quoted in Leon Cheinisse, *Les idées politiques des physiocrates* (Paris, 1914), 179.
1. *De l'ordre social*, 10.
2. *Ibid., 13*.
3. *Ibid.*, 23.
4. It is generally appreciated that Rousseau reasoned in this way, because of his strongly constructed case for subjectivism and for his pithy way of stating it: "What I feel is right, is right; what I feel is wrong, is wrong." The rationalists were not different from Rousseau in all respects. Their book of nature, like

this point well:

> The first principles of the social order [physiocracy, or the good economic society for Le Trosne] that reveal themselves to us are simple; they conform to the constitution of man; they have always been known implicitly. . . . They are incontestable, and their certitude is easy to know. They are established partly on moral notions generally admitted and expressly dictated by instinct (*par le sentiment intérieur*), and partly on the laws of nature itself, the results of which lie before our eyes. This social order is at the same time prescribed by justice and indicated by self interest; it has for its basis the rights and the duties of man as a moral and physical creature, of man fitted with intelligence and capable of discerning the true from the false, the good from the evil, the just from the injust, of man subject to [physical] needs and forced to make use of the means that the laws of reproduction offer him, laws which he discovers by experience and reflection.[5]

These ideas are independent of the world about the thinker. Le Mercier, for example, begins his classic exposition of physiocratic doctrine with an analysis of man — not too far removed from Descartes' *je pense, donc je suis* — and from it he deduces what the natural order should be, what man's rights and duties are, what political and economic laws flow from man's nature and from the physical laws regulating the universe. Le Mercier describes his method of reasoning accurately when he asserts: "I do not lay my eyes on any nation or any century in particular; I seek to describe things as they should be essentially, without bothering how they are or how they have been in any country whatsoever."[6] Again, he insists:

> As truth exists by itself and is the same in all places and all times, so by reasoning and examination we can arrive at it and all the practical consequences which result from it. Examples which appear to contradict these consequences prove nothing, for it is only that men have lost the way and do not have certitude and full knowledge of the truth.[7]

### III

The physiocratic theory of innate economic ideas should not be misunderstood and oversimplified on the basis of a few quotations like the above, as Locke had unjustly oversimplified Descartes' theory of innate ideas. The Physiocrats realized that babies were not born mumbling that a smaller supply means a higher price, or that agriculture is the source of all wealth. These things are innate, according to the Physiocrats, insofar as they are deduced from the

Quesnay's China, was pretty much a product of their own minds where they found "proof" of their principles.

5. *De l'ordre social*, 315–316.
6. *L'Ordre naturel*, I, 194.
7. *Ibid.*, I, 194–195.

first principles of knowledge with which all normal men are born. They are in the mind potentially at birth; they can be brought to the light of conscious knowledge by the right method, and without the intrusion of any information outside the mind itself.

The Physiocrats therefore sought to begin with incontestable first principles and to deduce from them, in typically rationalist, deductive fashion, their whole body of economic doctrine. Done in this method, they believed, their science would be foolproof. Here again Quesnay stands apart from his disciples. Except for him, however, the Physiocrats all subscribed to the deductive method of elucidating detailed laws of economic behavior from innate principles. Even Mirabeau, who was temperamentally incapable of being a rationalist, tried futilely to use the deductive method.[8] Le Mercier, however, was so successful in using it that his book is a rationalistic chain, each point being rigidly deduced from the preceding one, with the result that one weak link causes the whole argument to collapse. That is why contemporary critics condemned him for "his geometric step-by-step pace, his solemn stiffness, his abuse of the words *évident* and *nécessaire.*"[9] This is an indictment which puzzled the Physiocrats, for they constructed their chain of reasoning knowingly, in the full belief that it was the only correct method of procedure. Le Trosne, for example, introduces his *De l'intérêt social,* wherein he deals with specifically economic subjects, by telling the reader that he proposes to follow *"une logique exacte"* in deducing economic consequences from first principles stated in his *De l'ordre social.* "This theory," he continues in justification of his method, "constitutes an ensemble so united by a series of necessary and coherent deductions, that when the chain is broken one can present nothing more than scattered parts, isolated and disordered truths, parts of principles that are as obscure as the language is inexact."[1]

So the Physiocrats adopted the rationalistic system of deducing all truth from innate ideas. It was not a system into which they inadvertently fell, for they wanted to do for economic thought what Descartes had done for philosophy. They wanted to use his criterion

8. Mirabeau informs the reader of *L'Ami des hommes* that he wishes to begin with "general ideas" and "well-fixed definitions." His description of this work given in the last chapter indicates how poorly he followed the rationalist method of orderly deductive reasoning. There he says, quite correctly, that there is order only in the chapter titles, that the work is "inégal, sans goût, négligé, souvent diffus, et amphibologique . . . fatigue et étouffe." *L'Ami des hommes* (4th. ed.; Hambourg, 1758), 426.

9. The remark is Edgard Depitre's in his introduction to a later edition of Le Mercier's *L'Ordre naturel* (Paris, 1909), xviii.

1. *De l'intérêt social,* in *De l'ordre social,* 492–493.

of the clear and distinct idea for measuring the truth of each law they deduced, and they wanted to start, as he had done, with nothing except a few incontestable first principles. Then their conclusions would be irrefutable. Theirs would be the foolproof social science. One of the Physiocrats approvingly quotes the *Éloge de Descartes* by Thomas, in which the latter tells how the Cartesian system had been applied successfully to social thought as well as to philosophy. "This would be a grand project," the Physiocrat concludes, "to apply Descartes' doubt to these objects, to examine them one by one, as he examined all his ideas, and to judge everything only according to his great maxim of certitude."[2]

This is a procedure with which the other Physiocrats agree. All truths comprising the new science, Le Trosne insists, are to be decided as valid on the basis of *évidence*, "a sure and infallible guide,"[3] which he defines in Descartes' own terms: "a clear and distinct discernment of the ideas [*sentimens*] which we have, of all the perceptions which depend on them, and of all the relationships between these perceptions."[4] Le Mercier offers exactly the same definition of *évidence*, and then he goes on to insist that the new science accept as true only what is known with the certitude of *évidence*.[5] In this way, he believed, physiocracy would be demonstrable in the same way as was Euclid's geometry and it would hold sway over enlightened minds as sovereignly as Euclidian geometry then seemed to do.

Because they believed that men's minds naturally turned toward certitude and because they considered *évidence* irresistible,[6] all they asked for was liberty of discussion with which to expose their teachings. "It is of the very nature of truth to need only time for spreading itself and liberty for explaining itself. Its lawyer is *évidence* and its judge is reason."[7] It is mainly for this reason that the Physiocrats optimistically believed they had discovered a science and a method which would usher in the perfect society within a relatively short time.

2. "De l'utilité des discussions économiques," *Physiocratie*, IV, 46.
3. *De l'ordre social*, 305.
4. *Ibid.*, 306.
5. *L'Ordre naturel*, I, 84–86.
6. Dupont, for example, writes: "Il a vu qu'on ne pourroit résister à l'évidence et à l'autorité de ces lois souveraines quand elles seroient suffisamment connues et manifestées." "Discours de l'éditeur," *Physiocratie*, I, lv. And Le Mercier tells the reader: "Nos esprits ont une tendance naturelle vers l'évidence; et le doute est une situation importune et pénible pour nous. Aussi pouvons nous regarder l'évidence comme le repos de l'esprit; il y trouve une sorte de bien-être qui ressemble fort à celui que le repos physique procure à nos corps." *L'Ordre naturel*, I, 100.
7. "De l'utilité des discussions économiques, "*Physiocratie*, IV, 23–24.

This Cartesian certitude was a mathematical certitude which ruled out anything like prudential decisions for the economist. There was no longer to be any question of taking the better of two courses, no more weighing of advantages and disadvantages of alternatives. Things were right or they were wrong, and the clear and distinct idea separated the true from the false as mechanically as a grain grader separates little kernels from big. Descartes had applied mathematics to philosophy to create modern rationalism, and in the same way the Physiocrats had applied mathematics to social matters to come out with the exact science of physiocracy. It was a science, they insisted, which was subject in all respects to the laws of mathematical calculations, whether one dealt with the question of freedom of commerce, the "natural price" of grain, taxation, or wages. The laws of this science are clearly discerned, in Dupont's words, "by reflection, by judgment, by physical and moral arithmetic, by certain calculation."[8] Mirabeau, in the same way, looked upon Quesnay's *Tableau économique*, the basis of the new science in his opinion, as "the first rule of arithmetic, which was invented in order to reduce this elementary science [physiocracy] to an exact, precise calculus."[9] Le Trosne explains more fully how the new science is based on *"l'évidence mathématique"*:

> Calculus is a formula by which one works on measurable and comparable quantities, between which one seeks to discover some relationship. The result presents the unknown that was sought and which could be found only with a great deal of difficulty without this formula. Economic science, operating on measurable objects, is susceptible of being an exact science, of being subject to calculus.[1]

Such was the view the Physiocrats held of their new science, a mathematically exact science which properly treated of all man's social activity. Their writings, by and large, are consistent with their formal definitions of physiocracy. Le Mercier is probably the purest Cartesian rationalist of the school. Dupont and Baudeau rank close behind him. Le Trosne is undoubtedly a rationalist, but he uses historical examples and factual material to back up his arguments from time to time,[2] something Le Mercier disdains to do. Le Trosne's

8. "Discours de l'éditeur," *Physiocratie*, I, iii.
9. *Philosophie rurale* (Amsterdam, 1763), xix.
1. *De l'ordre social*, 320.
2. Le Trosne shows he has not completely disassociated his theory from the factual world about him by citing England as the great agricultural country of the time, Poland as the outstanding example of political disorder, and certain American colonies where he thinks one or another of the physiocratic principles is working out. He also appeals to history from time to time for additional proof of a point he seeks to make.

reasoning, however, stands independent of the factual material to which he alludes; it is brought in only by way of illustration or exemplification of the point he is making logically and deductively.

## IV

It has already been observed that the Physiocrats thought themselves to be nothing more than popularizers of Quesnay's discoveries. In their treatment of specific economic questions of the day, such as freedom of commerce, taxation, and the advantages of large farms over small, they adhere closely to the doctor's writings, and they occasionally invoke his authority. But when they deal with their concept of the new science they do not refer directly to his works or his conversation. As a matter of fact, Quesnay had written nothing on the subject as such. Students of economic thought have quite naturally — but incorrectly — assumed that the Physiocrats were accurate popularizers of their master's thought, and that he therefore agreed substantially with all their works. Auguste Oncken, the only student of physiocracy to think otherwise, observed late in the last century that he could not help believing that Quesnay was not "completely and consequently justly understood by any of his disciples. Various observations make us conclude that Quesnay, especially toward the end of his life, had the same feeling."[3]

It is therefore worth investigating whether Quesnay held the same concept of the new science of economics as did his followers. Since he never wrote on the subject itself, we must find his attitude toward economics by the indirect method of seeing how he handled economic questions, what method of reasoning he pursued, what proofs he used, how he arrived at conclusions. A word of caution is necessary at the beginning, if we would keep Quesnay's view of economics in proper perspective. The "new science" was just one phase — and a passing one at that — in Quesnay's intellectual life. He wrote his first economic article in 1756, when he was sixty-three years old and when he was already famous in France as a physician and a writer on both medical and philosophical subjects. His last article on economic matters appeared in February of 1768. After that time Quesnay showed no interest in physiocracy. He had transferred his speculation and his enthusiasm to mathematics, principally to intricate geometric problems. These last years were

---

3. *Op. cit.*, 721. Oncken does not enlarge upon this observation. The point is indicated in summary form, but not developed in Thomas P. Neill, "Quesnay and Physiocracy," *Journal of the History of Ideas*, IX (April 1948), 153–173.

a period of senile mathematical speculation — or at least so the Physiocrats thought.[4]

Quesnay's earlier writings throw some light on his ideas on the method proper to economic investigation. They reveal him as a typically well-rounded thinker of the eighteenth century, one who is saved from the accusation of dilettantism only by the sober good sense he always exhibited and by the relative profundity of his various essays. His nonphilosophical writings show him to be an independent thinker who stands above any one method or any system.[5] He lets the subject matter dictate the method he employs; he does not impose the same discipline on all subjects, as the true rationalist did. His articles "Évidence" and "Fermiers," for example, which appear in the same volume of the *Encyclopédie*, do not seem to have been written by the same man. They differ in method of reasoning, in style, even in temper. His essays on medical subjects, again, stress the value of observation and experience for arriving at new theories in medicine. His philosophical articles reveal a man versed in the rationalism of the Cartesian school, in Locke's empiricism, and in the older traditionalist philosophy of the Schoolmen.

Quesnay independently accepts or rejects various theories from each of these groups. In the main he agrees with Locke's stress on the role of property as a basis for society, for example, but he rejects Locke's empirical epistemology because he thinks it leaves the mind too passive. Again, he accepts many of Descartes' conclusions, but he decisively rejects his theory of innate ideas and his methodical doubt. In similar fashion, he shows independence in accepting much of Malebranche's reasoning about "order," but he rejects his occasionalist explanation for the interaction of the body and the soul. Thus the man who first wrote on economics in 1756 and was credited by his followers with discovering a "new science" was recognized as an eclectic who was competent to deal with medical, philosophical, and social subjects. His recognized competency in these fields is indicated by the articles he was assigned for inclusion in the famous *Encyclopédie:* "Évidence," "Fermiers," "Fonctions de l'âme," "Grains," "Hommes," "Impôts," "Intérêt de l'argent." He would likely have

4. In 1773 Quesnay published his *Recherches philosophiques sur l'évidence des vérités géométriques*, a work of quite low caliber mathematically that Turgot referred to as "scandal of scandals" [Gustave Schelle, *Du Pont de Nemours et l'école physiocratique* (Paris, 1888), 124.]

5. It is worth noting that the three *éloges* on which most information about Quesnay's life depends treat him as an eclectic philosopher who excelled in speculative medicine, wrote philosophical articles, and finally worked in agricultural economy. None of them looks upon him simply as the founder of physiocracy. All three *éloges* are contained in Oncken, *op. cit.*

received additional assignments if he had not severed relations with Diderot's publishing venture in 1759.[6]

The founder of physiocracy, then, was not a rationalist. In his article "Évidence," published in the *Encyclopédie* in 1756, he specifically and vigorously attacks the Cartesian theory of innate ideas subscribed to by the younger Physiocrats. "Innate ideas," he writes, "or ideas that the mind produces of itself without the action of any extrinsic cause, do not create in the·mind any proof of the reality of anything else."[7] Reason alone is not sufficient to arrive at a knowledge of the outside world and "natural laws" governing both man and the universe. Quesnay tells us in another place that "reason is to the mind what the eyes are to the body."[8] Knowledge sifts into the mind through reason and is acted on by it, but reason alone cannot procure knowledge of itself.

Consequently, Quesnay did not indulge in chain reasoning, as did Le Mercier and the rest of his disciples. In his economic articles, indeed, it is almost impossible to find any of the typically rationalist phrases, such as *évident* and *nécessairement*, or any of the "it-therefore-follows" transitions. Both Quesnay's background and his temperament militated against his succumbing to the rationalist form of reasoning, either in philosophy or in social science. By the time he began to write on economic subjects he was too old and too wise, too experienced and, as Baron Grimm put it, too "cynical"[9] to trust the Euclidian thinking which so intrigued his disciples.

That Quesnay did not look upon economics as a mathematically exact science is to be implied from his treatment of "natural law," the law on which his economics is based. His analysis of natural law shows clearly that he did not subscribe to the trend then current of converting the normative moral law into a Newtonian law of nature, as the other Physiocrats tended to do. He treats the subject

6. License to publish was withdrawn from the *Encyclopédie* in 1759 as a result of the commotion caused by Helvetius' *De l'esprit.* The government had revoked the *Encyclopédie's* license originally in 1752 as a result of the de Prades affair, but permission to publish was granted again in about a year. Because the *Encyclopédie* was always suspect and because Quesnay held a court position, the doctor could not continue his collaboration with Diderot after 1759. His three articles, "Évidence," "Fermiers," and "Grains," appeared under a pseudonym; "Fonctions de l'âme" should have appeared before "Grains," but for some reason Diderot never published it — mainly, this author believes, because of his disagreement with Quesnay's religious views expressed in the article. The other three articles were returned to Quesnay in 1759, at his request, and he gave them to Dupont for publication in the physiocratic journal.

7. In Oncken, *op. cit.,* 777.

8. "Droit naturel," in Oncken, *op. cit.,* 376.

9. *Op. cit.,* Pt. I, Vol. V, 481.

of natural law descriptively and analytically in two places: in his article "Droit naturel," first published in the September 1765, issue of the *Journal de l'agriculture, du commerce et des finances*,[1] and later included as the first article in Dupont's *Physiocratie;* and in the eighth chapter of his *Despotisme de la Chine*, wherein he tries to give a "systematic digest of the Chinese doctrine."

In both places Quesnay introduces his discussion of natural law by dividing it into physical laws and moral laws.[2] He then describes the former in terms which seem derived from Shaftesbury more than from anyone else. "By physical law is here meant the ordered course of all physical happenings most advantageous to the human race."[3] This is the law that the Creator has ordained for the operation of the universe, much as a designer may be said to have ordained the laws by which a jet-propulsion motor operates. It is the law which governs such things as agriculture, animal husbandry, the distribution of wealth, the operation of commerce and industry. When men follow this law faithfully they secure the greatest material benefits possible, as, for example, when they plant the right crops at the right time, use the right fertilizer and lay away the right amount for capital investment the following year.

Thus Quesnay optimistically believed that everything works properly for mankind's best material interests. All man need do is understand the law and work in harmony with it. This physical law is mathematically exact in its working out. Man is free to violate it, of course, but his doing so works out to his material disadvantage. These physical laws can be reduced to a science, Quesnay believed, just as Newton had collected and collated all the laws discovered by his predecessors to explain the working of the universe in his *Principia*.

Quesnay defines his second kind of natural law as "the rule of every human action in the moral order, conforming to the physical order evidently most advantageous to the human race."[4] These are

1. This was the first number of the *Journal* edited by Dupont, who, incidentally, was later fired for making it a physiocratic journal instead of the objective organ it was supposed to be. In his introduction to Quesnay's article on natural law, Dupont tells the reader that "this is the solid foundation on which the edifice [of economics] should rest." Oncken, *op. cit.*, 363. It is worth noting that in this introduction Dupont shows how Quesnay's distinction between physical and moral natural law made no impression on him. His first sentence is typical: "C'est la connaissance de l'ordre et des lois physiques et naturelles qui doit servir de base à la science économique." *Ibid.*, 362.

2. "Les lois naturelles sont ou physiques ou morales." "Droit naturel," Oncken, *op. cit.*, 374. The same distinction is made in the same words in the article on China. See *ibid.*, 637.

3. *Ibid.*, 375.

4. *Ibid.*

actions which had traditionally been considered "human" or "moral" or "ethical" acts, the dealings of man with man, such as the sale of a horse or the payment of a contract, acts which come under the heading of "commutative justice" with Quesnay. The rules governing such action are ordained by the Creator, he tells us, and man can promote his own prosperity, as well as that of mankind in general, by obeying these divine ordinances.

These two divisions of natural law, with Quesnay, have much in common. They are both discerned by the light of reason. They are both objective rules, absolutely independent of man's will, both ordained by the Creator for man's governance on earth. Their purpose is to promote man's happiness, which Quesnay seems to identify with his material welfare. They can both be disobeyed by man, for he has free will, but the penalty in each case is a material, physical punishment.[5] This "discovery" of the physical sanction attached to the moral law, indeed, was thought by his disciples to have been Quesnay's great discovery.[6] The fact remains, however, that Quesnay distinguished between the two kinds of natural law and he did not think them identical in all respects. The moral law dealt with what we can call "human actions." Nowhere does Quesnay try to reduce these actions to mathematical formulas, as he does the material comprised under physical law. He refused to believe that men would ever be sufficiently enlightened to follow their self-interest in such deterministic fashion as to surrender their freedom to make economically bad decisions — as the younger Physiocrats did. Nor did he believe that moral actions could be so "scientifically" classified as to rule out prudential decisions. Quesnay insisted on keeping man a free agent[7] — a point which his followers seemed anxious to overlook.

5. "Les transgressions des lois naturelles sont les causes les plus étendues et les plus ordinaires des maux physiques qui affligent les hommes." Oncken, *op. cit.*, 369. Quesnay stated in several places, however, that moral transgressions are visited with moral punishment as well. See his "De l'immortalité de l'âme," *ibid.*, 760–761.

6. It is true that the Physiocrats sometimes referred to the *Tableau économique* as Quesnay's great discovery. Dupont, Le Mercier, and Le Trosne all mention his discovery of a physical sanction attached to the moral law as a turning point in civilization.

7. "Il est de sa nature d'être libre et intelligent, quoiqu'il ne soit quelquefois ni l'un ni l'autre." "Droit naturel," Oncken, *op. cit.*, 370.

It is in his essay "La liberté," that Quesnay makes a thorough analysis of free will. On this point he is in almost perfect accord with the medieval Schoolmen. He indicates that free will is one of the two faculties distinguishing men from the lower animals. He then asserts that free will does not consist merely in freedom of choice between alternatives, but of a number of steps of the intellect from which a decision results.

V

Quesnay's method of arriving at economic theories is in keeping with his epistemology and his analysis of natural law. Any economic theorist of the mid-eighteenth century who rejected innate ideas and refused to subscribe to the deterministic concept of natural law then becoming popular in enlightened circles[8] would logically have to use the inductive method to arrive at economic laws, which, instead of being mathematical formulas, would be descriptive of tendencies. This is what Quesnay did. He used the inductive method knowingly, for he had shown his appreciation of this method of acquiring knowledge more than a decade before he published his first article on economics. In his preface to the *Mémoires* of the Royal Academy of Surgery (of which he was the secretary), Quesnay stated that "there are two sources for discovering truths which can enrich our art: observation and physical experiment."[9] In an essay on method which sounds surprisingly modern Quesnay goes on to tell how observation and experiment should be used as checks upon each other, how the observations of past generations and of a wide circle of contemporaries should be gathered for comparison, how research should never stop, and how practice and theory should be correlated for mutual support and mutual correction, since neither can be neglected without the other suffering.

Quesnay's articles on those economic subjects which involve man's free, human actions are also inductive in approach. In his article "Fermiers," for example, he starts out by warning the reader that "if one considers agriculture in France only under its general aspect, one can form only vague and imperfect ideas about it."[1] Therefore, he says, "consult the farmers themselves."[2] His article is a descriptive study of how farms are managed in the various parts of France, where horses are used instead of oxen, under what conditions it pays to use them, and how farming is conducted in England. From this description and comparative study of different farming techniques, Quesnay draws his conclusions. His article "Grains" is a similar kind of study, full of statistics and comparisons, in which he

8. Helvetius, for example, insisted in 1759 that "man is a machine" and that "the propositions of morality, politics, and metaphysics are as susceptible of demonstration as the propositions of geometry." Holbach reached the final conclusion of this trend in his *Système de la nature*, in which he claimed: "The same necessity which regulates the physical, also regulates the moral world, in which everything is in consequence submitted to fatality."

9. In Oncken, *op. cit.*, 724.

1. *Ibid.*, 159.

2. *Ibid.*, 160.

presents strong economic arguments for large farms rather than the *petite culture* which was one of the principal obstacles to an agricultural revolution in France similar to the one which had already made England the model farming country of Europe. Quesnay's writings are full of acute observations and of hardheaded realistic statements which stand in striking contrast to the deductive method advocated by his disciples. He shows himself aware that his reasoning is frequently in generalizations which have only limited value, as is the case with all abstractions. "We should not reason against the facts," he says in one place, "facts are realities. But a generic name, like the word 'commerce,' which confounds a multitude of different realities, is not itself a reality."[3] This is the very opposite of Le Mercier's advice to "ignore the facts."

There is one other aspect of Quesnay's economic thought which cannot be ignored: its mathematical aspect, seen for example in his *Tableau économique,* which so impressed his followers. Here Quesnay assumes that his perfect economic society has been established, and he shows how the annual product of the land would flow from the productive class through the hands of the proprietary and sterile classes back into the hands of the agricultural class again. This is one of the doctor's "economic problems" which deals with the working out of the physical division of natural law — and for which he showed a strong inclination.[4] It is, he wrote to Mirabeau, "a fundamental table of the economic order, [drawn up] in order to represent expenses and products in a way easy to understand, so as to judge clearly the order or the confusion that the government can cause [by its tax laws, its regulations on commerce and the like]."[5]

There is no doubt that Quesnay believed certain economic problems were subject to mathematical solution. And it is no less certain that he considered other economic subjects suitable only for descriptive and normative treatment. Using our more modern terminology, we can say that Quesnay treated some economic problems as properly parts of a social science and others as mathematical. He implicitly divided the subject matter of economics, then, into a physical or mathematical science on the one hand and a moral science on the other. This is a division which Quesnay never mentions explicitly, for he does not deal with the nature or the province of the

3. "Du commerce," in Oncken, *op. cit.,* 459.
4. Various people around the court, such as Marmontel, relate how Quesnay spent hour after hour tracing "zig-zag" lines on paper and working out mathematical problems. Ironically, his published tables and problems quite frequently contain simple mistakes of addition.
5. Quoted in Schelle, *Le Docteur Quesnay,* 389–390.

new science in any of his writings. And it is a division which his disciples never comprehended.

## VI

The younger Physiocrats took up the mathematical aspects of Quesnay's thought with even greater enthusiasm than their master. And they failed to maintain Quesnay's rounded outlook on economic questions, for, as Oncken has put it, they did not completely understand him. A comparison of Le Mercier's work with Quesnay's various writings clearly shows how the disciple produced an entirely different product than the master had in mind.[6] Le Mercier's analysis of liberty is typical. He treats liberty as one of the rights man enjoys in the natural order, and he begins his analysis of the subject by mentioning that he is unconcerned with the subject of "metaphysical" liberty, or freedom of the will, because the only liberty of importance is the social or physical liberty to do with one's property what one wishes. Quesnay, on the other hand, considered freedom of the will a basic point from which social liberty derived as a natural right. So, whereas Quesnay grounded liberty in man's nature, Le Mercier based it on property rights. It is not too much to say, we believe, that none of Quesnay's disciples understood his philosophical writings — on which his economic thought was ultimately based.

Quesnay seems to have realized this sometime after the appearance of Le Mercier's *L'Ordre naturel* in 1767. He quit writing for the *Éphémérides du citoyen*, and his contacts with his disciples became less and less frequent. Similarly, his disciples seem to have realized that a gulf was widening between their master and themselves, a fact which they probably put down to his advancing age. Whether Quesnay was growing senile by 1768, as may well have been the case, is beside the point. It is with his previous writings that the Physiocrats dissented, whether they were willing to face the fact or not. The fate of Quesnay's then unpublished manuscripts indicates that his disciples realized their master's economic thought was not as clear-cut and rationally arranged as their own. Quesnay had placed in Dupont's hands those articles which he had withdrawn from the *Encyclopédie* when it fell under government censure. Dupont was

6. In 1788 Mirabeau wrote to his friend Longo: "I have seen the author of *L'Ordre naturel et essential des sociétés politiques* work six whole weeks in the doctor's apartment writing and rewriting his work — and then deny his father and his mother." Since all the Physiocrats insisted Le Mercier interpreted Quesnay, we can interpret this statement only as meaning that the interpretation by Le Mercier was not a faithful one.

supposed to publish these articles, three of which dealt with economic subjects. The one on interest he published in January, 1766, but the school was embarrassed by the fact that it did not fit into their rigid theory. They always passed this article by in silence when discussing interest. The other two articles, "Hommes," and "Impôts," were not published until early this century.[7] Although Dupont had over five years in which to publish them, he never let them appear in print. There seems no adequate reason for this omission except his belief that they were not good physiocracy.

In that respect, Dupont was right. For it was the old story, repeated with the Physiocrats as with so many other schools of thought, of the master not being faithfully followed by his disciples. If Karl Marx could protest to his son-in-law, "I am no Marxist," Quesnay could claim with greater vehemence that he was no Physiocrat. The doctor had written in an age of transition. He was an older man who knew the traditional thought coming from the Schoolmen which was taught in the Sorbonne until the middle of the eighteenth century. He was not immune to newer influences, but he was independent in selecting some items and some methods from them while rejecting others. Thus in his treatment of natural law Quesnay offers a mixture of old attitudes and new. He looks on the law as being moral, but at the same time he seeks to add the modern element of a physical sanction to its observance. Quesnay never thought his position through to ultimate conclusions, nor did he ever push back to ultimate presuppositions. For that reason he never created a system, properly speaking. His doctrines were a mixture of ideas and theories which he handled with consummate common sense. It remained for his followers, particularly Le Mercier, to reduce those theories to a system, subject them to a single method, and place them upon frankly recognized basic assumptions.[8]

We are inclined to think that the younger Physiocrats would have privately nodded in agreement if they had read the contrasting estimate made of Le Mercier and Quesnay by those collaborating Encyclopedists, Diderot and Grimm. Of Le Mercier's exposition of physiocratic doctrine Diderot observed: "I admire the certitude and the fertility of his principles, the easy manner in which he solves the

7. "Hommes" was published in 1908 in the first number of the *Revue d'histoire des doctrines économiques et sociales*, edited by Étienne Bauer. "Impôts" was published later in the same volume, edited by Gustave Schelle.

8. In 1939 Max Beer published his *Inquiry into Physiocracy*, in which he tried to establish the essential similarity of physiocracy and the economic teaching of the Schoolmen. It is worth noting that Beer used Quesnay's writings almost exclusively to establish his point.

gravest difficulties, and the simplicity with which he resolves objections."[9]     Of Quesnay Grimm wrote: "He is not only naturally obscure, he is even systematically obscure and he pretends that truth should never be stated clearly."[1]  Quesnay had not succumbed to the rationalistic deductive method of building the new science of economics.  His disciples had.

THOMAS P. NEILL.

ST. LOUIS UNIVERSITY

9. *Oeuvres complètes de Diderot* (Paris, 1875), XVIII, 273.
1. *Op. cit.*, 481.

I

In Chapter XIX of *The New Economics*,[2] Professor Leontief examines what he considers to be the fundamental postulates of Keynes' *General Theory*. As a modern classicist, he is concerned with evaluating the Keynesian assumptions and their effect upon classical doctrine. He imputes to Keynes two postulates which, he alleges, Keynes substitutes for the single underlying assumption of the traditional theory. The Keynesian postulates concern the supply of labor and the demand for money, respectively.[3] Professor Leontief asserts that the Keynesian theory is not more but less "general" than the classical theory, because the two postulates in question replace — in his view — the single classical assumption regarding economic choice, which, of course, if true would be contrary to the principle of parsimony.

According to Professor Leontief, Keynes assumes, (1) that money wages are rigid downward, and (2) that the rate of interest is insensitive to an increase in the quantity of money and is also, therefore, rigid downward. I wish to question the assertion that Keynes assumes, as a basic postulate, a downward rigidity in money wages; though, on the other hand, I wish to assert that the phenomenon actually occurs. I shall agree that Keynes does assume, as a basic postulate, a downward rigidity in the rate of interest; but, I also propose to question Professor Leontief's interpretation of liquidity preference and to restore the Keynesian theory of the demand for money to its proper station in economic theory. *In fine*, I must add that even if the *General Theory* should be shown to rest upon a greater number of assumptions than the classical system, the principle of parsimony is a monitory rather than a prohibitory commandment.[4]

1. I wish to acknowledge the assistance gained from innumerable conversations with Professor Richard S. Howey, University of Kansas, and to thank Doctor James Tobin, Harvard University, who read the manuscript and made many helpful suggestions. Responsibility for the views expressed in the paper is, of course, my own.

2. A collection of Keynesian literature, edited by Seymour E. Harris (New York: Alfred A. Knopf, 1947).

3. "The nature of the supply of labor and that of the demand for money are the two principle points of divergence between the basic postulates of the *General Theory* and the teachings of the classical doctrine." *Op. cit.*, p. 233.

4. Though it is not the purpose of this note to discuss the relative "generality" of the two systems, the following observations appear to be pertinent. Manifestly, there is no definition of "generality" generally accepted by economists.

Although Professor Leontief deals first with the supply of labor and then with the demand for money, his argument would proceed more logically in the opposite direction. That is to say, he would be more methodical were he first to imply the fallaciousness of the theory of liquidity preference, then to insert the postulate of rigid wages in order to enable the Keynesian model to be at underemployment equilibrium, and finally to occupy himself with a destructive analysis of the latter assumption as it appears in his special context. However, I shall follow the order employed by Professor Leontief.

## II

To let Professor Leontief introduce his case, I may quote: "Traditional analysis considers the aggregate quantity of labor supplied, in the case where this supply is a competitive one, to be a function of the *real* wage rate; Keynes on the contrary assumes that up to a certain point — defined by him as the point of full employment — one particular level of *money* wages exists at which the supply of labor is perfectly elastic and below which no labor can be hired at all. The deliberate exclusion of the cost of living as a determinant of labor supply makes the latter independent of the level of *real* wages."[5]

The first Keynesian postulate then, according to Professor Leontief, is that wages are rigid downward.[6] But Keynes at no point in his

Professor Leontief measures "generality" in terms of fundamental postulates, Keynes in terms of employment equilibria.

Professor Leontief does perform a great service in distinguishing between postulates and deducible conclusions when he reprimands those of Keynes' disciples who maintain that the classicists assume unemployment to be non-existent.

5. *Op. cit.*, p. 233. The downward rigidity in money wages is, it is true, indispensable to Keynes' definition of involuntary unemployment. I am not concerned in this paper, however, with the role that wage rigidity plays in this context. Rather, I am interested in examining the contention of neoclassicists to the effect that the Keynesian system requires such a rigidity in order to be at underemployment equilibrium. Dr. Lawrence R. Klein, in his admirable tract on the nature of *The Keynesian Revolution* (New York: Macmillan, 1947), has clearly relieved the Keynesian model of any theoretical dependence upon the definition of involuntary unemployment which appears in *The General Theory*. Cf. *The Keynesian Revolution*, pp. 80–90.

6. Cf. also *The New Economics*, Chapter XIV, "The General Theory," by G. Haberler. Professor Haberler apparently agrees with Professor Leontief. Quoting from p. 167 in the Haberler chapter: "Keynes assumes that (money) wages are rigid downward. If this assumption, which is certainly not entirely unrealistic, is rigidly adhered to, most of his conclusions follow: Under-employment equilibrium is radically changed. Obviously, under-employment equilibrium with flexible wages is impossible — wages and prices must then fall continuously, which can hardly occur without further consequences and cannot well be described as an equilibrium position."

writings explicitly states such an assumption.[7] Professor Leontief does not say that Keynes makes an explicit statement to that effect; he imputes the postulate to Keynes in what he considers to be a logical clarification of Keynes' thesis. I believe the supposed logical necessity for the postulate, which seems obvious to Professor Leontief, to be invalid; and I shall attempt to show that it is a result, in Professor Leontief's mind, of his misinterpretation of the other matter of liquidity preference; that, in short, if this misinterpretation is corrected, any logical necessity for a postulate of wage rigidity disappears.

Parenthetically, I wish to account for the space which the subject of wage rigidity occupies in *The General Theory*. Keynes devoted

Manifestly, Professor Haberler is tying the Keynesian argument to the postulate of rigid wages. He continues in a similar vein on p. 169: " . . . it is pointed out (in Chapter 19 of the *General Theory*) that a reduction in money wages will usually influence employment, but in an indirect fashion, through its repercussions upon the propensity to consume, efficiency of capital, or the rate of interest. The last mentioned route, via the interest rate, is the one most thoroughly explored by Keynes and the Keynesians. As wages and prices are allowed to fall, money is released from the transactions sphere, interest rates fall, and full employment is eventually restored by a stimulation of investment. This amounts to giving up the idea of under-employment equilibrium under a regime of flexible prices and wages except in two limiting cases: Full employment may be prevented from being reached via this route, (a) if the liquidity trap prevents a fall in the rate of interest . . . or (b) if investment is quite insensitive to a fall in the interest rate. Keynes himself regarded both these situations not as actually existing but as future possibilities. But what if we do regard them as actually existing — which as a short-run proposition, allowing for dynamic disturbances through unfavorable expectations, etc., would be by no means absurd? We would still not have established a stable under-employment equilibrium, for wages and prices would still continue to fall."

There seems to be some difficulty here with the meaning of equilibrium. I should presume that by an equilibrium at less than full employment, Keynes simply meant the condition where there are no forces inherent in the economic system which tend to raise the level of employment to that of full employment. This is precisely the condition which exists when the liquidity trap halts the decline of the rate of interest, or when the schedule of the marginal efficiency of capital has an interest elasticity of zero. That is to say, in either of these cases, there would be nothing further to gain from a reduction in money wages. It would follow, perforce, that the unemployed would not, theoretically at least, continue to offer themselves at lower money wage rates. If the unemployed blindly did so, the fact that prices and wages would continue to fall, approaching zero as a limit, is irrelevant. For there is nothing in Keynes' definition of under-employment equilibrium that precludes such a downward spiral of prices. The point is that in these two situations, the economist would no longer be justified in advising the labor unions to accept lower money wages. It follows that in these limiting cases cited by Professor Haberler we have conditions of under-employment equilibria *without* the assumption of rigid wages. It is the rigidity in the rate of interest or the perfectly inelastic schedule of the marginal inefficiency of capital which brings about the condition of equilibrium.

7. Cf. *op. cit.* Chapter XI, "The General Theory," by Abba P. Lerner, reprinted from the *International Labour Review*, October 1936.

many pages to the subject, but he probably did so because he wrote with one eye cocked in the direction of the counsels of state. In other words, Keynes felt it would be politically inexpedient to ask the labor unions to accept lower money wages. In the first place, he realized that it would be foolhardy to expect a single labor union to accept lower money wages; for it would thereby suffer a loss in real wages to the benefit of the rest of the economy. In the second place, Keynes was, if anything, not a socialist; and he undoubtedly abhorred the thought of a wage deflation at the behest of an authoritarian government. It was as a result of his personal predilections, therefore, and of political realities in the Great Britain of 1935, that Keynes concentrated his attention on alternative avenues to full employment.

However, Keynes not only rejected the classical prescription for full employment on practical grounds, but for theoretical reasons as well. He knew that if labor unions voluntarily accepted lower money wages, full employment would not necessarily follow.[8] Keynes reasoned as follows: Given a condition of less than full employment, if unemployed workers offer themselves at a wage less than the going rate, wages and marginal costs would necessarily fall. But since incomes would fall, *pari passu*, with the decline in the marginal cost of labor, there would be a commensurate decrease in demand, and entrepreneurs would have no reason to increase total employment over the given underemployment level.[9]

Keynes was aware, of course, of the fact that a reduction in money wages might have the indirect effect of increasing employment via the rate of interest. To follow the argument employed by Keynesians and modern classicists alike,[1] as incomes fall, less money is needed for transactions, hoards increase, the rate of interest falls, and investment and employment rise. However, I have already noted the impracticability of such a policy. Furthermore, it takes

8. On this account Keynes considered the workers in their intuitive apprehension, to be superior as economists to the theorists who embrace the classical doctrine. Cf. *The General Theory*, p. 14.

9. Similar reasoning, of course, applies to the correlative but equally fallacious argument sometimes made by labor leaders: an appeal for higher wages on the ground that an increase in wages would increase effective demand and employment.

To deal sensibly with the relationship between a change in the wage rate and the level of employment, a general equilibrium model is required. The dichotomy of price and distribution theory according to the traditional approach, with the oversimplification implicit in economic analysis which artificially separates the producers' and consumers' market, necessarily yields a meaningless solution.

1. Cf. note 6 above; also *Lapses from Full Employment*, (London: Macmillan, 1945) for the Pigovian version.

no great imagination to visualize the damage probably sustained by the schedule of the marginal efficiency of capital were such a program to be embarked upon. And furthermore, to clinch the Keynesian argument from a theoretical standpoint, the road to full employment may be blocked by the nature of the demand for money.

I must postpone until later a consideration of the Leontief version of the demand for money. It is sufficient to say at this point that the postulate of wage rigidity is a gratuitous one. Keynes was motivated by considerations other than theoretical in calling our attention to wage rigidities. Moreover, as has been suggested, I shall show that this postulate is not only not explicit but, in addition, that it is not implicit in the Keynesian system.

Surprisingly enough, after attributing the assumption of rigid wages to Keynes, Professor Leontief proceeds to deny its empirical validity.[2] But ironically, excepting, of course, the case of the total-itarian state, the existence of wage rigidity may be asserted.

. If money wages are rigid downward — and a cursory survey of labor management relations will show that they are — the worker's behavior may not be completely consistent with the assumption regarding economic choice which underlies classical economic theory. In this respect, I believe Keynes' insight into the worker's psychology was superior to the hedonistic calculus which is fundamental in the traditional doctrine. Even though we treat all workers as *economic men* or more correctly, as human beings endowed with as much irrationality as the rest of us, I cannot accept a supply of labor theory cast in the classical mold as empirically sound. In the first place, a bargaining worker always hopes that his cut of the "economic pie" will be larger if he obtains a higher money wage.[3] In the second place, even though the worker be forewarned that his realization of higher money wages will leave him no better off, he (and certainly I should share his irrationality in a comparable situation) would be irrational enough to choose the higher absolute level of wages and prices. This would be the case, I believe, if for no other reason than that the worker would *feel* better off with the inflated pay check. These facts of worker psychology, which will not be denied by many, are overlooked by Professor Leontief.

In the following paragraphs he describes and criticizes a Key-

2. Professor Haberler evidently would not agree with Professor Leontief on this point. Cf. note 6, above.

3. Moreover, a higher money wage received by a single labor union will mean higher real wages for that segment of labor. To this extent, the money illusion is not a manifestation of irrationality, but rather of rationality on the part of the worker.

nesian utility function. "In contrast to the classical, this Keynesian utility function would include, among the ultimate constituents of an individual's preference varieties, not only the physical quantities of (future and present) commodities and services but also the money prices of at least some of them. In particular the *money* wage rate would be considered as entering directly the worker's utility function: confronted with a choice between two or more situations in both of which his real income and his real effort are the same, but in one of which both the money wage rates (and, consequently, also the prices of consumers' goods) are higher than in the other, he would show a definite preference for the former. A classical *homo economicus* would find neither of the two alternatives to be more attractive than the other.

"From such a monetary utility function, a monetary supply curve of labor can be easily derived. In contrast to its classical counterpart, it will show the labor supply as dependent not only on the relative but also on the absolute prices and wage rates . . . .

"Although neat and internally consistent, such 'psychological' interpretations of the monetary element of the Keynesian theory of wages are hardly appropriate. They contradict the common sense of economic behavior. The reference to the fact that no worker has ever been seen bargaining for real wages — even if true — is obviously beside the point, since while bargaining in terms of dollars the worker, as any one else, can still be guided in his behavior by the real purchasing power of his income. Moreover, the 'psychological' interpretation of the monetary element in consumers' behavior deprives Keynes' unemployment concept of its principal attribute. Why should any given rate of employment or unemployment be called 'involuntary' if it is determined through conscious preference for higher money wages as against larger real income?"[4]

4. *The New Economics*, pp. 235–6. In this passage, Professor Leontief reiterates a stand on *The General Theory* taken earlier in his article, "The Fundamental Assumption of Mr. Keynes' Monetary Theory of Unemployment," this *Journal*, November 1936, pp. 192–197. At that time, Professor Leontief imputed only one basic postulate to the Keynesian system: that pertaining to the supply of labor. It was Professor Leontief's contention that the fundamental difference between Keynes and the classicists lay in the former's refusal to accept the latter's *homogeneity postulate* as applicable to the supply of labor.

If we let $z$ represent the supply of labor, $y$ the wage rate, and $x_1, x_2, \ldots x_n$ the prices of wage goods, the classical supply function may then be expressed as

$$z = f(y, x_1, x_2, \ldots x_n)$$
$$f(ty, tx_1, tx_2, \ldots tx_n) = t^b f(y, x_1, x_2, \ldots x_n)$$
$$\text{where } b = 0$$

That is, the classical supply function is homogeneous of the zero degree.

The underlying implication in these paragraphs to the effect that laborers calculate according to strict economic choice is simply not warranted by the facts of economic experience. It is precisely this kind of inflexible thinking on the part of classical economic theorists that renders economic theory so vulnerable to attack by specialists

The reference made by Professor Leontief to the *homogeneity postulate* is simply an econometrician's way of saying that the money illusion of *The General Theory* amounted to a serious departure from the fundamental assumption regarding economic choice of classical economic theory. In a later rejoinder, "The General Theory of Employment," this *Journal*, February 1937, pp. 209–223, Keynes agreed that this difference between the two systems does exist. (For an evaluation of the relative importance of this difference, see *The Keynesian Revolution*, p. 83.)

Assuming that the validity of the money illusion, or rigidity in money wages, is accepted, this same money illusion, Professor Leontief argues, must logically be applied to all supply and demand functions in the Keynesian system. The same view is expressed by Dr. James Tobin in his article on "Money Wage Rates and Employment," *The New Economics*, pp. 572–587 — a definitive treatment of the nasty problem of tracing the effect of changes in money wage rates on the level of employment. To quote Dr. Tobin: "Without the retention of the 'homogeneity postulate' for all supply and demand functions except the labor supply function, the Keynesian money wage doctrine cannot be maintained . . . . When the existence of variable factors other than labor is admitted, Keynesian theory requires that these factors be fully employed and that their prices be perfectly flexible. This is where the 'homogeneity postulate' — the assumption of 'rational' behavior — enters with respect to the supply functions of these factors. If the sellers of these factors were, like the sellers of labor, influenced by the 'money illusion', their prices would be rigid like wages and there could be unemployment of these factors. A change in the money wage rate could then alter the employment of labor by causing a substitution between labor and other factors."

"Keynes, since he assumes away the existence of other factors, presents no reason for this distinction between labor and other factors. Lerner, however, asserts that it is 'plausible and in conformity with the assumption of rationality of entrepreneurs and capital owners, who would rather get something for the use of their property than let it be idle, while labor has non-rational money-wage demands.' " (*op. cit.*, p. 582). Dr. Tobin, however, takes issue with Professor Lerner and argues that it is just as reasonable to expect the sellers of these other factors to have nonrational money price demands as wage earners.

Similarly, with the consumption function, Dr. Tobin argues: "The Keynesian consumption function, which is crucial to the Keynesian solution to the money wage problem, is framed in real terms . . . . This is the application of the 'homogeneity postulate' to the consumption function. If 'money illusion' occurred in consumption and saving decisions, real consumption expenditure would depend on the level of money income as well as on the level of real income . . . . A change in the money wage rate, changing the level of money incomes and prices, would alter the real demand for consumption goods and therefore, affect the volume of both output and employment. Here again, therefore, retention of the 'homogeneity postulate' is an essential assumption for Keynesian money wage doctrine.

"But if wage-earners are victims of a 'money illusion' when they act as sellers of labor, why should they be expected to become 'rational' when they come into the market as consumers?" (*op. cit.* p. 583.)

Although it is my purpose in this paper to show that the Keynesian system is not dependent, fundamentally, upon the wage rigidity, it is interesting and

in other social sciences. If the economic theorist aspires to a position of influence in public policy he must sufficiently complicate his theory so as to give proper recognition to the irrational elements of society's economic behavior. This is not the place to dwell at greater length on the shortcomings of classical economic theory with respect to

useful to deal with the problem posed by the countless rigidities which, of course, exist in the real world. As for the Tobin application of the money illusion to supply functions other than labor, it seems obvious that the homogeneity postulate correctly describes the supply of the fixed plant and equipment. Once the entrepreneur is saddled with fixed costs he will operate as long as he can defray variable costs. Rigidities in the prices of variable factors other than labor are irrelevant, since if labor is to be substituted, it will be substituted for plant and equipment, not for raw materials or goods in process.

But suppose we admit, what I do not believe to be true, that the suppliers of capital equipment are beset by the money illusion, and if you please, in the short run! Even in this situation, can anyone soberly maintain that, say, ten millions of unemployed laborers could be brought into employment by accepting lower money wages? Anyone making such a plea would have to show that the elasticity of the demand for labor is sufficiently greater than unity to make up for the decrease in expenditures on capital equipment in order to maintain the level of effective demand. To my knowledge, the discovery of this fact remains for the economist of the future. Surely, the burden of proof lies on the shoulders of him who makes the plea.

In any event, in a world of imperfect markets and rigidities in price policies, the entrepreneur is more sensitive to changes in effective demand than to changes in marginal costs. Even if wages fall, the entrepreneur is not likely to increase employment in the short run unless effective demand increases.

Of course, in a sense, all unemployment is voluntary. That is, full employment could always be achieved by reducing money wages sufficiently to cause a reversion to a handicraft society. But such a solution to the problem of mass unemployment implies the untenable assumption of a change in tastes involving a transference of demand from the highly fabricated goods of the roundabout process of production to goods which can be produced by a handicraft society. Needless to say, the solution of the problem of unemployment by means of a "Molly McGuire" movement on a mass scale would be intolerable.

As for the Tobin consumption function, it seems to me that more is being made out of the money illusion than should be. All that the money illusion means when applied to the suppliers of labor is that they associate a higher real wage with a higher money wage. In reality, this does not amount to irrational behavior, but rather rational behavior; because an individual worker or labor union that obtains an increase in money wages will find its real wages have risen above what they would have been had the increase in money wages not been achieved, at least momentarily.

It is of interest to note, as Dr. Tobin has done, that, abstracting from expectations, consumers' real wealth might be increased, possibly leading to increased consumption expenditures, by a general deflation of wages and prices. (Cf. *op. cit.* p. 584. Also, James Tobin, "The Fallacies of Lord Keynes' General Theory: Comment," this *Journal*, November 1948, pp. 763–770, esp. p. 769.) This is the 'Pigou effect,' investigated by Professor Don Patinkin in his recent article, "Price Flexibility and Full Employment." (*American Economic Review*, September 1948, pp. 543–564.) In other words, abstracting from expectations, a reduction in prices may have a stimulatory effect via the propensity to consume as a consequence of an appreciation in the value of assets held in the form of cash."

modern psychology. It is enough to underscore the superior insight evinced by Keynes and revealed so clearly in a comparison of the Keynesian theory of the supply of labor with that of the classicists.

It must be admitted that management and labor alike are hardly unconcerned about real wages, though they deal in terms of monetary units. This is not to say, on the other hand, that the worker has not usually found his real income rising, at least temporarily, with an increase in money wages. Consequently, he is primarily concerned with money, rather than real, wages. As for the perhaps unnecessarily obscure Keynesian definition of unemployment, the Keynesian worker obviously associates a higher real income with a higher money income, whether his expectations are fulfilled or not. It is elementary that the breadwinner is more conscious of the monetary value of his take-home pay than of its real value, subject, as the latter is, to the erratic movements of the prices of thousands of commodities.[5]

Professor Leontief is doubtless correct in criticizing Keynes' use of the word, "involuntary," when he asks: "Why should any given rate of employment or unemployment be called 'involuntary,' if it is determined through conscious preference for higher money wages as against larger real income?"[6] However, I should parry with the question: "Has Professor Leontief, or anyone else, ever seen a worker confronted by such a choice?" But on with the Leontief case.

In order to make *The General Theory* more palatable, Professor Leontief posits a minimum wage law in deference to the Keynesian wage rigidity.[7] With the minimum wage law, Professor Leontief is on familiar ground. All that is necessary now, says our modern classicist, is to inflate all prices except wages. However, Professor Leontief must dispose of one more obstacle, liquidity preference, which, if unexorcized, would render nugatory the classicist's prescription of inflation. The discussion must turn, therefore, to the second basic postulate, which pertains to the demand for money, and which I must investigate to some extent before completing my evaluation of Professor Leontief's treatment of the Keynesian supply of labor.

### III

Professor Leontief is, I believe, correct in assigning liquidity preference an important role in the Keynesian system. I feel, how-

5. It is true that labor unions are presently paying closer attention to changes in price indices than formerly. Whether or not this fact will eliminate the downward stickiness in wages remains to be seen.

6. Cf. footnote 4, p. 559 f.

7. This he does while at the same time scoring Keynes' reluctance to clothe his wage rigidity in institutional raiment (*op. cit.* p. 239). A case in point is Keynes' lengthy digression on labor union strategy noted above.

ever, that certain aspects of the Leontief interpretation merit further reflection. In Professor Leontief's eyes, liquidity preference is nothing more than the velocity of the circulation of money in new, if less fashionable, dress. Before stating my most serious objections to the Leontief interpretation of Keynes, I wish to examine his confusion of the velocity of the circulation of money with the demand for money.[8] In Keynes, all money is divided between that used to satisfy the transactions motive on one hand and the speculative and precautionary motives on the other hand, $M_1$ and $M_2$ respectively. We have:

$$M = M_1 + M_2$$
$$M_1 = T\ (Y)$$
$$M_2 = L\ (r)$$

The demand for money for transactions is then a function of income; whereas, the demand for money as an asset is a function of the rate of interest.[9] The $V$ of the quantity equation, to the contrary, does not provide us with such useful tools of analysis, for it is the velocity of circulation of all money in the system. It is in this confusion of $V$ with $M_2$ that we shall find the key to the error in the Leontief interpretation.

According to Professor Leontief, "The theory of liquidity preference provides the Keynesian system with a deflationary mechanism which defeats, through the process of automatic hoarding, every tendency toward inflationary reduction of involuntary unemployment."[1] In this statement Professor Leontief betrays his classical bias and preoccupation with inflation as a prescription for full employment.

It will serve my purpose to consider the Leontief interpretation in two parts, depending on the position of the rate of interest relatively to the liquidity preference schedule. Let us first assume the rate of interest to be above the level where it becomes insensitive to the quantity of $M_2$, i.e., above the point where the interest elasticity of the liquidity preference schedule approaches infinity. In this range, it is certainly correct, to quote Professor Leontief again, that "The theory of liquidity preference provides the Keynesian system with a deflationary mechanism which defeats . . . every tendency toward *inflationary* reduction of involuntary unemployment."[2] True, in

8. Cf. *The New Economics*, Chapter XLI, "Keynes and Traditional Theory," by R. F. Harrod, p. 603.

9. For a statistical investigation of the shape of the "L" function see James Tobin, "Liquidity Preference and Monetary Policy," *Review of Economic Statistics*, May 1947, pp. 130–1.

1. *The New Economics*, p. 238.

2. Cf. preceding footnote. Italics mine.

Keynes we do *not* reach full employment through inflation.[3] However, this is not to say that Keynes denies that the prescription of increasing the quantity of money achieves full employment. It is only with regard to the mechanism by which an increase in the quantity of money affects the price level that Keynes parts company with the classicists. And here the divergence between the two systems is clear!

Let us grant for purposes of the argument the wage rigidity of the Leontief interpretation. The classicist recommends an increase in the quantity of money. Then, according to the classicists, people will find in their possession a quantity of cash greater than their accustomed holdings. In trying to rid themselves of the excess, they will force up the price level, Professor Leontief's supply and demand curves for labor are raised, and involuntary unemployment is wiped out.[4]

It is at this point that my earlier reference to the Keynesian treatment of money becomes significant. Starting again with the assumption of unemployment and rigid wages, the Keynesian agrees to an increase in the quantity of money. But, contrary to the classical belief, this expanded money supply will *not* result in inflation. Let us suppose the central bank expands the quantity of money in circulation by purchasing securities on the open market. Former security holders find their cash holdings increased above their accustomed level. Do they then make haste to the consumers' market to exchange their excess holdings of cash for goods? Of course not![5] They try to get

3. By *inflation*, I mean a rise in the money prices of commodities. Hence, an increase in the quantity of money would not entail inflation unless accompanied by an increased demand for commodities, or unless there is full employment and the interest rate falls, thereby causing an increase in the rate of investment. Abstracting from frictions, such increased investment could take place without causing inflation if there were unemployed resources in the economy.

4. *Op. cit.*, p. 236.

5. Cf. *op. cit.*, Chapter XIV, p. 170. Professor Haberler apparently shares, to some extent at least, Professor Leontief's views on this subject. Quoting Professor Haberler: . . . "we must assume, it seems to me, that consumption is not only a function of income but also of wealth (and *liquid* wealth in particular)." (Italics mine.)

Strictly, consumption (that is, the absolute magnitude of expenditures on consumer goods) is some function of income. We have:

$$C = F\,(Y)$$

The shape of the consumption function is in turn determined by a host of other variables, which we would normally take as parameters, such as wealth, the tax structure, the social security system, expectations regarding the future availability of consumer goods, the stock of consumer durables on hand, price expectations, government policy (e.g. "E" Bond campaigns), advertising, the status of thrift as a social virtue, the extent of national self-sufficiency, the ratio of dividends to corporation profits, the relation between urban and rural population, the extent

rid of their excess money assets by bidding against each other and the central bank for securities, thereby driving the prices of securities up and the interest rate down to a level where they are content to hold the increment to the money supply. Investment increases, and equilibrium at full employment is obtained without inflation.

The point is that in the classical system the holders of the increment to the money supply treat this increment as income. The fallacy in this presumption is clearly revealed in the Keynesian analysis. This increment of money is treated as an asset, not in the general accounting sense, but as an asset in the Keynesian sense of money held to satisfy the speculative or precautionary motives. The $V$ of the quantity theory equation, then, does not provide us with a microscope of sufficient power to detect the fundamental workings of the economy. The superiority of the Keynesian instrument should be obvious.

I shall turn briefly to the situation wherein the rate of interest is rigid downward, i.e., the case in which the interest elasticity of the liquidity preference schedule approaches infinity, and an increase in the quantity of money has no effect on the rate of interest. This case is sometimes referred to as the *liquidity trap* or *two per cent case*. It

---

of education, expectations as to the future level of incomes, and perhaps (cf Harold Lubbell, "Effects of Redistribution of Income on Consumers' Expenditures," *American Economic Review*, March 1947, pp. 157–170) the distribution of income. All of these factors combine to determine the public's attitude toward saving or consumption at various levels of national income. Professor Haberler evidently wishes to include in this list of variables which determine the shape of the consumption function, not only wealth, but also some variable reflecting the proportion of wealth held in liquid form. He apparently feels that the post-war inflationary trend in prices was due in part to holdings of wealth in the form of cash and in the quasi-liquid form of government bonds. (Cf. "Causes and Cures of Inflation," by Gottfried Haberler, *Review of Economics and Statistics*, February 1948, pp. 10–14). It seems more plausible to me that the post-war upward shift in the consumption function be explained by the dearth of consumer durables sustained by the war, plus expectations of further price rises. In other words, it does not seem necessary to concede that people have suddenly decided to convert their assets into consumer goods in order to explain the upward movement of the consumption function following in the wake of war.

Of course, if monetary policy is successful in stimulating new investment, incomes and consumption will rise. And if the economy is at or approaching full employment, prices will rise. But this price inflation is due, directly, to the increment to income flow and only indirectly to the increase in the quantity of money. In this event, and particularly if open market operations are accompanied by additional income stemming from deficit financing on the part of the government, the marginal propensity to consume may approach unity due to price expectations, with a runaway inflation in the offing. It would surely be agreed, however, that such a state of affairs would be abnormal, reminiscent, e.g. of the catastrophic German monetary experiment following World War I, and need not detain us further at this juncture.

is evidently this *trap* which Professor Leontief has in mind when he refers to the " . . . *deflationary mechanism which defeats, through the process of automatic hoarding,* every tendency toward inflationary reduction of involuntary unemployment."[6]

If I understand him correctly, Professor Leontief denies the very existence of the *liquidity trap.* Evidently, in his opinion, people hold money to buy commodities. If the prices of commodities rise or fall, people hold more or less money, respectively, in order to keep the purchasing power of their holdings constant. If this be the case, liquidity preference cannot vitiate the inflationary process, since individuals would be willing to increase their holdings of cash only as the price level rises. In other words, if the quantity of money is increased, people will try to rid themselves of the increment of cash. In so doing, they will force prices up to a level at which they are willing to hold the increment of money. That is, prices will rise to a point where the purchasing power of their cash holdings is the same as it was before the quantity of money was increased.

Here again the unsurmountable obstacles which modern classicists necessarily encounter when they attempt to reduce Keynes to classical stature become apparent. The Leontief delineation of the Keynesian system is inaccurate, as he refuses to pay close attention to the respective roles which $M_1 = T \ (Y)$ and $M_2 = L \ (r)$ play in the Keynesian system. He insists on defining liquidity preference in terms of the $k$ of the Cambridge cash balance version of the quantity theory.[7] In Keynes, of course, the increment of money is not treated as income. The investor finds the prices of securities, land, housing, *et. al.,* too high. Consequently, he adds to his holdings of cash anticipating an eventual upswing in the rate of interest, thereby nullifying the effect of an increase in the quantity of money on the rate of interest.[8] Hence, this is one[9] of the Keynesian *cases* in which

6. Cf. note 1, p. 563. Italics mine. It should be pointed out that Keynes envisaged this rigidity in the rate of interest as only a remote, but probable eventuality. Despite this fact, Professor Leontief apparently devotes his entire attention to this case.

7. Where all $M = kY$ and $k$ is defined as the constant proportion of their incomes which individuals hold measured in real terms.

8. It is true that in a purely static model the demand for money would be proportional to prices. But abstraction from uncertainty is unthinkable. Hence, Professor Leontief suggests (*op. cit.,* pp. 238–40) that Keynes should have written *The General Theory* in dynamic rather than comparative static form. Certainly, it would be exceedingly helpful to have at hand a dynamized, general equilibrium model encompassing all important economic relationships. However, the dearth of dynamic economics, combined with data showing the relatively large number of students in the field of economics, clearly indicate something more than the dereliction of economists as the source of the difficulty.

9. Cf. footnote 6, p. 555 for the Haberler version of other *Keynesian cases.*

fiscal policy must be substituted for monetary policy if equilibrium at full employment is to be attained.

## IV

It has been my purpose to question the Leontief interpretation of *The General Theory* on the following grounds. First, Keynes realized that rigid money wages must be taken into account by the practicing economist. However, he did not postulate wage rigidity, for it is not essential to his theoretical structure. Secondly, full employment may be attained through monetary policy which does not entail a rise in the prices of commodities. Thirdly, among other reasons, we may have an unemployment equilibrium because of the nature of the demand for money. Finally, I wish to reiterate that we were in equilibrium at less than full employment in the *liquidity trap case* due to a rigidity in the rate of interest, not because of a rigidity in wages.[1] Obviously, the wage rigidity plays such an important role in the Leontief version of Keynes only because of the former's conception of liquidity preference and his denial of the *liquidity trap*.

IRA O. SCOTT, JR.

HARVARD UNIVERSITY

## COMMENT

Mr. Scott's comments on the Keynesian theory of wages bring to my mind the story about the woman who when accused of having returned a borrowed crock in damaged condition defended herself by claiming that first of all she had not borrowed the crock; secondly, the crock was already broken when she borrowed it; and thirdly, that it has been returned in perfect condition. After admitting, as Keynes did himself in his answer to my earlier reference to that particular point, that the "money illusion" plays a prominent part in *The General Theory of Employment*, and conceding that without it the very concept of involuntary unemployment becomes meaningless, Scott denies that it constitutes an indispensable element of the Keynesian argument. His reference to Lawrence Klein's success in

1. In *The Keynesian Revolution*, p .89, Dr. Klein concludes: "In order to show that full employment is not automatic in a perfect world subject to the Keynesian conditions, it is necessary to assume nothing whatsoever about rigidities in the system, but only to make plausible assumptions about the interest-elasticity of certain basic relationships." In the *case* of the perfectly inelastic marginal efficiency of capital, no rigidity is involved. In the *liquidity trap case*, on the other hand, there is a downward rigidity in the rate of interest.

*"relieving*[2] the Keynesian model of any theoretical dependence upon
the definition of involuntary unemployment" hardly makes my critic's
position any more consistent.

Considering the singularly elusive character of many of his
theoretical formulations and the impressionistic nature of some of
Keynes' writing, I feel that simple quotations from *The General
Theory* will be of little use in settling the present controversy. Let
it be noted however that James Tobin, whose article on "Money
Wage Rates and Employment" is recommended by Scott as a defin-
itive treatment of the problem, flatly states that "clearly one of
Keynes' basic assumptions . . . is that 'money illusion' occurs in the
labor supply functions."[3]

Turning now to his discussion of liquidity preference, I still do
not see that Scott succeeds in shaking my fundamental contention
that Keynesian theory represents an ingenious but not very successful
attempt at treating essentially dynamic problems in static terms.
He does not contradict my observation that the relationship between
the quantity of money and the rate of interest is due entirely to the
operation of the speculative motive. On the contrary, his insistence
on what I cannot but consider to be nothing more than a typical
pedagogical simplification — the conceptual division of all available
money into two *additively* combined parts — $M_1$ and $M_2$ — only
serves to emphasize that particular point. Unfortunately, he does
not say "yes" or "no" to my contention that under stationary con-
ditions, (i.e., a situation in which none of the relevant economic
variables such as the rate of interest, prices, etc., are changing or are
expected to change over time) nobody would hold any money for
speculative purposes; that is, that in such a situation $M_2$ would be
*identically* zero. But this is the crux of the whole matter. If the
foregoing observation is correct, then in a truly stationary situation,
the cornerstone of the Keynesian theory of interest — the direct
relationship between the quantity of money and the rate of interest
— falls to the ground. ($M_1$ is assumed not to be a function of $p$.)

Mr. Scott's plausible description of the sequence of events which
would follow an expansion of the money supply by the central bank
might be right or wrong; insofar as it refers to an obviously dynamic
process, it does not and cannot contribute to the clarification of the
theoretical issues involved in this particular controversy. Had he
tried to analyze this loosely depicted process in terms of an explicitly
formulated dynamic model, he would have found that (a) either the

2. The italics are mine.
3. *The New Economics*, p. 580.

original or the final position of his system must prove *not* to be a truly stationary equilibrium position, or (b) the existence of speculative money holdings, $M_2$ under purely stationary conditions would have to be assumed.

In the two per cent or the "liquidity trap" case, Mr. Scott explicitly refers to anticipation of an eventual upswing in the rate of interest as the reason why the investors are ready to absorb and immobilize any additional supply of money. This clearly is a dynamic phenomenon. Under conditions of truly stationary equilibrium, characterized by constancy over time of the magnitudes of all relevant variables including the two per cent rate of interest, there would be no reason to expect its "eventual upswing" and to hold on to additional cash for speculative purposes. To justify the existence of a liquidity trap as a long run stationary phenomenon, one would have to give up any reference to expected changes and fall back, as in the case of the labor supply, on some kind of a money illusion, i.e., a demand for money assumed to be independent of its purchasing power. This is essentially the position explicitly taken by Don Patinkin in the article quoted by Scott.[4]

<div align="right">WASSILY LEONTIEF.</div>

HARVARD UNIVERSITY

## FURTHER COMMENT

In note 6, p. 555, Mr. Scott tries to refute my contention that the existence of underemployment *equilibrium* depends on the assumption of rigid wages and that underemployment equilibrium is impossible, if there is competition among wage earners. My reason was that competition would drive down *money* wages and prices so long as there is involuntary unemployment — a situation incompatible with an equilibrium position. It should be noted that real wages need not go down. In fact within the simplified frame of the Keynesian system where prices are supposed to fall parallel with wages, real wages would not fall at all. Mr. Scott holds that if the liquidity trap prevents a fall in the rate of interest or if the elasticity of demand for capital with respect to the rate of interest is zero, "the unemployed would not, theoretically at least, continue to offer themselves at lower money wage-rates" and "the economist would no longer be

4. Passing from the defense of the Keynesian position to a counterattack, Patinkin presents in a later article, "The Indeterminacy of Absolute Prices in Classical Economic Theory," (*Econometrica*, XVII, No. 1, January 1949) what he considers to be a proof of internal inconsistency of the classical position. I am dealing with his arguments in a note in the January 1950 issue of *Econometrica*.

justified in advising the labor unions to accept lower money wages."
(By the way, who are the economists who advise unions to accept
lower money wages? It is surely not those economists to whose
advice unions are at all likely to listen.)

Taken literally, Mr. Scott's argument moves in a curious kind
of circle: Economists would not be justified in recommending wage
reductions and the unemployed would not be willing to work at lower
wages, because there "would be nothing further to be gained from a
reduction in money wages," meaning obviously that employment
would not increase. In other words, my theory is wrong, because
economists and the unemployed have convinced themselves that it is
wrong and prevent it from being tested by keeping wages rigid!

It is hardly necessary to say that this kind of argument does not
prove anything. It does not even come to grips with the issues
involved. But let me point out the obvious fact that if the unemployed
"do not offer themselves at lower money wage-rates," they are either
not involuntarily unemployed or there is no competition in the labor
market.

That Keynes' theory of underemployment equilibrium depends
on the assumption of rigid wages should by now be clear to everybody.
I feel embarrassed by saying the same things over and over again,
but since again and again there appear articles which overlook or
ignore the most obvious considerations, I feel constrained to repeat
a few things which I have said already in the first edition of my
*Prosperity and Depression*[5] and repeated in later editions and else-
where.[6]

The proposition that there can be any amount of unemployment
with rigid wages is entirely compatible with classical and neoclassical
economics. On the other hand, if there is competition in the labor
market and elsewhere[7] and if, hence, wages and prices fall continu-
ously or at least spasmodically so long as there is involunatry un-
employment, the real value of the quantity of money increases without
limit and, assuming that the liquidity preference curve is perfectly
elastic or the marginal efficiency entirely inelastic, it is clear that
something in the Keynesian system has to give way, for instance, the

5. Chapter X, §9 (in later editions, Chapter XI, §9).
6. *Op. cit.*, Chapter I, §5, and Chapter XIII, §6, and the essay quoted by
Mr. Scott.
7. There is no space here to explain in detail that it is not necessary that
there should be *perfect* competition. Some flexibility of wages which is quite com-
patible with considerable deviations from perfect competition would be sufficient
for the argument.

propensity to consume. What Don Patinkin[8] somewhat pompously enunciates as "the Pigou effect," viz. the proposition that the consumption function will shift upward, if the real value of liquid assets (or assets in general) reaches a certain level, is just common sense which must have been in the minds of many writers. At any rate, when I stated it in the first edition of my *Prosperity and Depression*, I thought that I expressed an obvious fact. Furthermore, it is only one possibility. For example, exactly the same argument that holds for the propensity to consume applies also to the marginal efficiency of capital.

It is true, of course, that all this is not the whole story and that it does not settle all problems of policy. Dynamic repercussions upon the marginal efficiency of capital, changes in the income distribution due to the increased real value of money debts and similar frictions are vastly more important than the mechanics of the static Keynesian system. Surely Keynes would have been the first to acknowledge that. It is high time that Keynesians recognize the inability of the Keynesian system to cope with the problem of wage and price flexibility so that the discussion can be moved out of the dead-end track on which it has been shunted by those writers who take as the last word the static Keynesian system which its author, if he had lived longer and had not been preoccupied with other matters, would have abandoned long ago as an untenable intermediate station.

GOTTFRIED HABERLER.

HARVARD UNIVERSITY

8. "Price Flexibility and Full Employment," *American Economic Review*, September 1949, *passim*.

# RESOURCE MALALLOCATION WITHIN FIRMS

The literature of welfare economics is replete with instances of disparities between social and private product. The monopoly problem, as traditionally conceived, was a special instance of such a discrepancy, it being held that monopolies invested too few resources in production as compared with firms producing for purely competitive markets. The advent of the theory of monopolistic competition, with its systematic introduction of such additional variables as selling effort and product quality, has suggested the possibility that resources may be malallocated, not only *among* firms, but *within* firms. For example, a firm may invest too much or too little of the economy's resources in product quality, relative to its investment in output; thus too much quality and too little output, or vice versa, will result.[1]

For the sake of analysis, let us imagine a single firm, producing and selling a branded product for which there are no perfect substitutes. The management of this "monopolistic competitor" is concerned with determining the quality and output of the product which will maximize static profits.[2] We shall suppose that quality and output are each single and unequivocal dimensions. Advertising, selling expense, and all manifestations of promotion are disregarded in the sense that they are impounded in *ceteris paribus*. The firm charges a price, uniform to all buyers, which just sells the current output.[3] Under these assumed circumstances, the management should strive to keep the ratio of marginal revenue to marginal cost on account of output $(\Delta R_o/\Delta C_o)$ equal to the ratio of marginal revenue to marginal cost on account of quality $(\Delta R_q/\Delta C_q)$, and both equal to unity.[4]

1. These ideas, together with much of what follows, were originally developed by the author in 1942 while writing a doctoral dissertation under the supervision of Professor Chamberlin; subsequently, the press of events has postponed their restatement in article form. Some of the other writers who have considered this same subject include: R. M. Shone, "Selling Costs," *Review of Economic Studies* (2), June 1935; H. Smith, "Advertising Costs and Equilibrium," *Review of Economic Studies* (2), October 1934; R. F. Kahn, "Some Notes on Ideal Output," *Economic Journal*, March 1935; and, of course, E. H. Chamberlin in *The Theory of Monopolistic Competition.*

2. The following discussion is timeless in that the slow acquisition of good will or ill will, the possibility of changing tastes, incomes, and degrees of competition, or the fact that maximum profits over time may require temporary price restraint, etc., are all ignored.

3. The rather simple model employed here is very similar to the one used in "Profit Maximization under Monopolistic Competition," *American Economic Review*, June 1941.

4. Symbolically, total revenue $(R)$ is a function of output $(O)$ and quality $(Q)$ and of nothing else; total cost $(C)$ is similarly determined exclusively by the management's decisions regarding $O$ and $Q$.

If the writer were a better draftsman, it might be possible to develop the argument more readily with the aid of a three dimensional figure, the horizontal axis representing output, the axis in depth standing for quality, and the vertical axis depicting total revenue and total use value.[5] A more pedestrian procedure is to dispense with the axis in depth, representing product quality, and instead to employ the dollar versus output diagram used here; then, when we alter the product quality, the $U$ and $R$ schedules will shift up and down in our figure.[6] It will be demonstrated first, that $\Delta U_o/\Delta R_o > 1$ (which is quite familiar); second, that $\Delta U_q/\Delta R_q \gtrless 1$ (which is not so well understood); and third, that, apart from coincidences, managerial policies will hence cause $\dfrac{\Delta U_o}{\Delta C_o} \gtrless \dfrac{\Delta U_q}{\Delta C_q}$ (which is most disturbing from the viewpoint of welfare economics).[7]

(1) *The relation between $\Delta U_o$ and $\Delta R_o$.* Let us suppose that the firm's management decides to invest extra resources in expanding the rate of output from $O_1$ to $O_2$. The quality of the product is such as to give the total use value schedule $U_1$ in the figure. The schedule $R_1$ shows the total revenue obtained from the sale of different outputs, at nondiscriminatory prices, when the original quality, upon which $U_1$ is based, is maintained. The total revenue obtained from sales was $DO_1$ before the increase in output[8] and $EO_2$ after it.[9] In this particular case the diagram has been drawn so that $\Delta R_o$, which is $EO_2 - DO_1$, is zero. However $\Delta U_o$, as shown by the rise in $U_1$ between $A$ at output $O_1$ and $B$ at output $O_2$, is $BC$. So long as it is

5. If the product in question were a consumer good, the total use value would correspond to the money value of the total utility obtained by the buyers. The difference between total use value ($U$) and $R$ is consumers' surplus. The value of $U$ is, of course, a function of $O$ and $Q$ and is also expressed in monetary units; it corresponds to the total revenue which might be obtained if a given output were sold, unit by unit, by a perfectly discriminating monopolist.

6. Just as a demand schedule shifts when buyers' incomes change.

7. Welfare economists could rest more comfortably if there were reasons for supposing that free enterprise usually resulted in an equality between $\dfrac{\Delta U_o}{\Delta C_o}$ and $\dfrac{\Delta U_q}{\Delta C_q}$ at profit maximizing equilibrium. (See p. 575 below.)

8. Because $OD$ (the slope of which represents the selling price) is constructed parallel to $HA$, the tangent to $U_1$ at $A$ (the slope of which represents the marginal use value). The rate at which buyers purchase is such that their marginal use value is equal to the firm's selling price. The quantity so determined, multiplied by the price, is total revenue ($R$). Geometrically, at each and every vertical section, the slope of $U_1$ must equal that of a line drawn to $R_1$ from the origin.

9. Because $OE$ is constructed parallel to the slope of the tangent to $U_1$ at $B$.

true that use value rises less than proportionately with increases in physical quantity, which is generally considered axiomatic in economic theory, $\Delta U_o > \Delta R_o$, and there is a buyers' surplus.
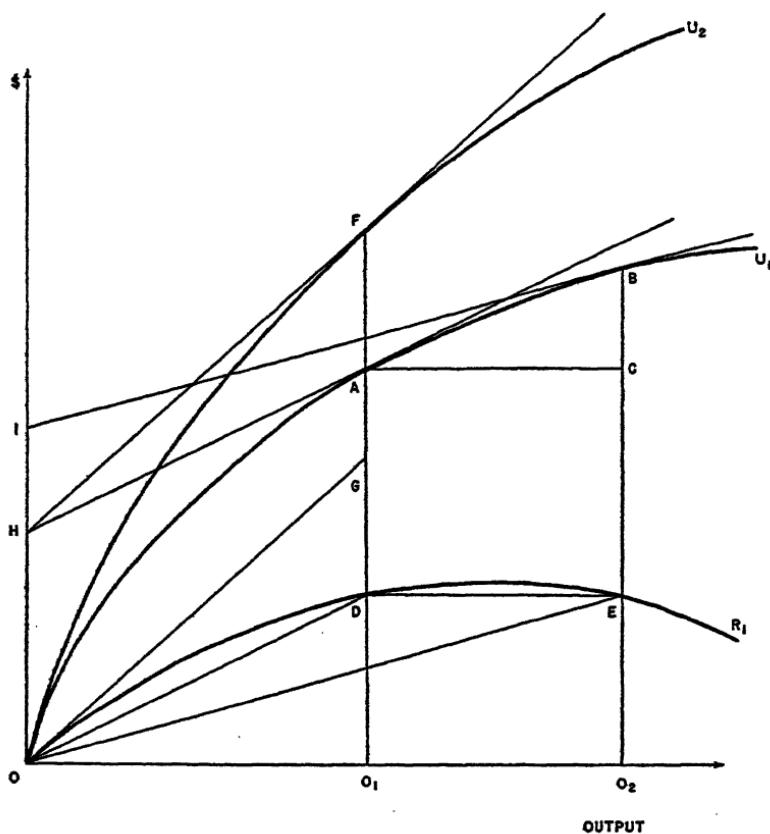


OUTPUT

(2) *The relation between $\Delta U_q$ and $\Delta R_q$.* Let us suppose that, instead of increasing output, the firm now produces a new and improved quality of product. A new total use value schedule, against output, is now needed, and this is $U_2$. The increase in the total use value at $O_1$ quantity, due to the improved quality, is represented by $FA$.

By drawing through $O$ a line $OG$, parallel to the tangent to $U_2$ at $F$, we obtain the increment to revenue, $GD$. In the figure, the curve $U_2$ has been drawn so that the slopes of $U_1$ and $U_2$ at $A$ and $F$ respectively make a common intercept on the vertical, or dollar, axis at $H$; hence $GD = FA$, or in other words $\Delta U_q = \Delta R_q$. However,

this is evidently a special case, since $U_2$ may in general have various slopes at $F$, with the result that the tangents at $F$ and $A$ may have dissimilar intercepts on the dollar axis, even though the value of $\Delta U_q$ remains $FA$; thus, in general, $GD \gtreqless FA$, or in other words, $\Delta U_q \gtreqless \Delta R_q$.

(3) *The socially significant relation between $\Delta U_o/\Delta C_o$ and $\Delta U_q/\Delta C_q$.* It is a commonplace theorem that profit maximization requires an equating of marginal revenue and marginal cost in the various dimensions. In the present case we can specifically suppose that management will see to it that $\Delta R_o$ equals $\Delta C_o$, that $\Delta U_o > \Delta R_o$ because of the tendency for the marginal utility of a good to decline, and hence that $\Delta U_o > \Delta C_o$. However, $\Delta U_q \gtreqless \Delta C_q$, because, while management will equate $\Delta R_q$ to $\Delta C_q$, $\Delta U_q \gtreqless \Delta R_q$. Hence it will only be by sheer accident that $\Delta U_o/\Delta C_o = \Delta U_q/\Delta C_q$. Thus, in the abstract, one can expect a malallocation of resources, from the social viewpoint, within this firm.

The basic explanation of this tendency for resources to be misdirected within the firm is that the consumer can usually decide the quantity he wishes to purchase whereas the quality is already determined by the firm.

This statement is naturally subject to many reservations. The consumer often finds that his quantity adjustment is also far from being a continuous affair; for example, when buying automobiles, which in terms of his budget are "sizeable" items. Also the consumer may be able to select from a number of competing brands having different quality characteristics.

However, it is still true that the quality adjustment of the consumer, accomplished by selecting one of several rival brands, will often be more discontinuous than his quantity adjustment. One of the essential ideas of the theory of monopolistic competition is that differentiated products are imperfect substitutes. When the "closest" substitute is "far away," and altogether too imperfect, the consumer will not seek, through buying other brands, to escape the quality determination of the firm from which he buys. When monopoly elements are strong, and substitutes are far removed, there is a graver danger that profit maximizing may occasion too little or too much quality, relative to the firm's investment of resources in output.[1]

1. Incidentally, product quality may not be a continuous variable for the manufacturer either, but be a discreet possibility instead; then there is no third

It does not follow, however, that economic welfare would be advanced if a wider choice of quality were available. Such a result might lessen intra-firm misdirection of resources of the kind considered here, but the economic cost of providing such a wide choice of qualities, (especially as there are so often increasing returns to scale in manufacturing) might render this "solution" undesirable.[2] One must also distinguish between the number of choices (which may be optimum) and what free enterprise decides they shall be (which may not be optimum.)[3]

The preceding analysis has supposed that $\Delta U$, $\Delta R$, and $\Delta C$, whether relating to output or quality, are always ascertainable. We know that management in practice often fails to maximize profits because these marginal values cannot be estimated accurately. A government agency which tried to equate ratios of marginal use value to marginal cost would have the same trouble. The practical significance of the fact that management supposedly has one goal (i.e., $\Delta R_o/\Delta C_o = \Delta R_q/\Delta C_q$) while government might have another goal (i.e., $\Delta U_o/\Delta C_o = \Delta U_q/\Delta C_q$) is lessened to the extent that both management and government are more likely to miss than hit their targets.

Notwithstanding these reservations, there are theoretical grounds for supposing that managerial attempts to maximize profits may lead to a malallocation of resources within the firm. Except by chance, too many or too few resources may be devoted to quality, to promotion, or to output. Here is another instance of divergence between private and social product to be added to the long list of doubts and qualifications relating to untrammeled laissez faire.

STEPHEN ENKE.

UNIVERSITY OF CALIFORNIA, LOS ANGELES

and invisible axis in depth, and a two dimensional diagram such as the figure emphasized here is alone permissible.

2. Cf. E. H. Chamberlin's discussion of a "sort of ideal" compromise between the benefits of quality choice and the cost of providing it in terms of reduced output efficiency (*op. cit.*, p. 94).

3. Some of the ideas developed by A. P. Lerner and H. W. Singer ("Some Notes on Duopoly and Spatial Competition," *Journal of Political Economy*, XLV, 176–182) are applicable by analogy to this problem.

# PSYCHOLOGICAL PRODUCTION AND CONSERVATION

## I. PSYCHOLOGICAL PRODUCTION

In the study of economic processes, students of the subject have concentrated most heavily upon increasing the total economic value of scarce goods through production, defining the latter as changing the form, place, ownership, and time of use of a good. It seems to me that viewing production in this limited sense places an unnecessary restriction upon economic investigation. The viewpoint taken here is that economics should concern itself with the study of all factors which influence the total economic value of scarce goods, and the psychological factor is one of these which is often overlooked. Consequently, those processes which are designed to control a person's desire for goods and, therefore, their total economic value must be viewed as productive or destructive.

We might expand our concept of production by saying that any activity is productive which increases a person's desire for a good that he owns, or which he is about to own, through a change in his tastes. Contrariwise, an activity which decreases his desire for such a good would be a destructive activity. Very few economists have regarded this type of production to be within the realm of economics. The basic utility of a good because of one's subjective value is often relegated to the disciplines of moral philosophy and psychology. However, the matter is of such importance in economics that it cannot be dismissed so readily from the discipline. An objective methodology in economic investigation would permit of no preference in considering the means whereby total economic value or satisfaction may be increased or decreased. It is as important to determine the influence of changes in the quantity and quality of desires on the total economic satisfaction derived from goods as it is to determine the influence of the traditional changes in form, place, ownership, and time of use of goods. Not to do so makes the economist guilty of arbitrariness and subjectivism, and diminishes correspondingly the scientific character of economics.

Traditional preoccupation with market price has focused attention upon man's desires as they contribute to the strength of "demand." In the study of exchange, "desire" is viewed statically and primarily with respect to its influence upon the pricing of goods which are in the process of sale.[1] However, it actually is *after* goods have been

1. As the psychological force underlying man's productive activities, desire has been viewed as both static and dynamic. It is with respect to acquired goods that dynamic desires have been largely ignored.

purchased that they are put to the economic test, and only then will it be discovered how much satisfaction they will give.  It is at this latter time that "desire" plays its most fundamental role, a role which has for the most part been slighted.  The satisfaction derived from accumulated wealth deserves as much attention as the prospect of satisfaction to be derived from future production, sales, and purchases.

A mathematical formula for total subjective economic satisfaction or value might be stated as follows: $S = \dfrac{G_a \, Q_a \, A}{G_d \, Q_d \, I}$

The symbols in the equation have the following meanings:

$S$ = total personal economic satisfaction

$G_a$ = the aggregate of scarce goods acquired

$Q_a$ = a quality index of the scarce goods acquired

$A$ = the degree of appreciation which a person holds for the quantities and qualities of goods acquired

$G_d$ = the aggregate of scarce goods desired, including those already acquired as well as those not yet acquired

$Q_d$ = a quality index of the scarce goods desired

$I$ = the intensity with which the quantities and qualities of the scarce goods are desired

The "law" which the foregoing equation represents may be stated as follows: *A person's total economic satisfaction varies directly with the quantity and quality of the scarce goods which he acquires, and directly with the degree of appreciation which he has for them; it varies inversely with the quantity and quality of the goods he desires, and inversely with the intensity of those desires.*

Undoubtedly the above equation presents many difficulties, perhaps insuperable, if one would attempt to assign quantitative values to the various factors in an effort to arrive at some kind of numerical value for $S$.  These difficulties are unavoidable inasmuch as $Q_a$ and $Q_d$ involve the qualities of goods, which are certainly difficult to rate.  It is particularly with respect to $A$ and $I$ that measurement is perhaps impossible in view of the fully subjective nature of these factors.  Nevertheless, despite the difficulties which might make it impossible to arrive at significant mathematical measures of $S$, the algebraic equation is useful because it is a concise way of showing the relationships which exist between the various factors that influence total personal satisfaction and the relationships of these factors to total personal satisfaction itself.[2]

2. Similar difficulties, though not so great, attend the mathematical application of the equation of exchange. The latter is primarily useful in expressing in mathematical form the relationship between factors affecting price levels.

Economic satisfaction is a state of psychic equilibrium or repose between a person's desires for economic goods and the gratification which he receives from the goods he acquires or has acquired. In terms of the above equation, this psychic equilibrium for a given person might be thought of as involving a balance between the various elements, such that $S = 1$. However, for the vast majority of persons we find that individual economic desires exceed considerably the gratification received from acquiring goods or from those already acquired. Thus, symbolically, $G_dQ_dI > G_aQ_aA$ for the average person. It follows then that for most persons $S < 1$.

Emotional balance with respect to economic desires and their gratification may be achieved in one or more of six basic ways:

1. Acquiring more goods.
2. Acquiring goods of a higher quality.
3. Developing a greater appreciation for goods already acquired and being acquired.
4. Developing a desire for fewer goods.
5. Developing a desire for goods of lesser quality.
6. Developing a less intensive desire for the quantities or qualities of goods that a person does not have.

Points 4 and 5 actually embody point 6 carried to the ultimate degree. Thus, when the intensity of a person's desire for a given quantity and quality is reduced to zero, he no longer desires that particular quantity and quality of a good.

In this note, I should like to call attention particularly to points 3, 4, 5, and 6, which indicate the dynamic role which "desire" itself plays in arriving at economic satisfaction. The object of economic activity is neither always nor solely that of acquiring goods (points 1 and 2) to meet given desires. Desires themselves are active determinants of total personal economic satisfaction. Consequently, it is very much within the province of economics to investigate social actions with the purpose of determining the extent to which they are economic. The more that social actions are in accord with points 3, 4, 5, and 6 (as well as points 1 and 2) the more economical they are insofar as they would increase personal economic satisfaction.

The basic economic law mentioned above points out that if we assume a given number of scarce goods to remain constant the satisfaction which the user receives varies inversely with the quantitative and qualitative standards of his desires. Controlling desires for the purpose of attaining economic satisfaction involves four out of six possible methods, and it is probably true that they are as important

in meeting the economic problem for the average person as is the method of acquiring scarce goods. A good becomes more useful, or its utility is increased, if a person learns to appreciate it more than formerly. Moreover, a person's economic dissatisfaction is reduced if he decreases his desires to a point where they involve a quantity and quality of goods that he can actually acquire. This type of control may be both the cause and effect of greater appreciation of the goods he has already acquired. Controlling desires in such a way as to promote greater economic satisfaction by increasing the economic value of certain goods may be termed psychological production.

In view of the foregoing, it would be indefensible arbitrariness to think that the purpose of economic activity is merely the production of more and more goods regardless of whether man's satisfactions are increased thereby. The assumption that overcoming scarcity invariably leads to economic satisfaction or that it is the only effective approach to economic satisfaction is untenable. It may well be doubted whether the total economic satisfaction which a person of high income receives from his wealth is necessarily greater than that which a person of moderate income derives from his. Indeed, of the two, the person of moderate income may receive even greater economic satisfaction (psychic equilibrium) if his wealth very nearly equals the standards which make up his desires. On the other hand, the person of high income may have desires for economic goods which have far outgrown his greater income. Overcoming the scarcity of goods alone does not insure the attainment of a solution to the economic problem. It is equally important that the quantitative and qualitative level of a person's desires be not too far above his actual ability to acquire scarce goods. This is often the more realistic approach for the average person in achieving economic satisfaction.

An interesting relationship exists between $G_a$ and $A$ by virtue of the operation of the law of diminishing marginal utility. Thus, as the wealthy person amasses a larger and larger quantity of economic goods, each additional unit yields a decreasing amount of satisfaction. Should this amassing continue, marginal utility might evidently become zero or even negative. But it is not necessary to go that far before a person realizes that the possession of many economic goods requires care and often produces anxiety. The maintenance and use of two automobiles demand more distracting attention than is demanded by one. A larger house and grounds bring forth more problems than a small house on an average sized lot. A larger staff of servants increases the burden of household management. The same holds true with the amount of entertaining of guests and the

elaborateness of such entertainment. It is also a commonplace that persons possessing many valuables have more to fear from theft and robbery than persons with few, and that they may often fear for their lives and those of their families because of their wealth. Such considerations, by diminishing $A$, would increase the possibility that very wealthy persons might possess less psychic equilibrium and thus less economic satisfaction than the person of moderate means.

While the importance of controlling desires in gaining economic satisfaction must be recognized, it does not follow, as some might facetiously rejoin, that we might as well forget about acquiring more goods and concentrate upon a determined and thorough curtailment of all or most of our desires for scarce goods. That would be as one-sided an approach as for a person to acquire excessive varieties and quantities of goods without coming closer to his ever-expanding desires. Economic satisfaction can most easily be attained when there is some compromise between acquiring more and better goods (especially the essentials for health, comfort, and decency) and scaling down unreasonable desires. Since it is difficult and sometimes impossible to reduce desires to the point of sacrificing essentials, it is agreed that production in its traditional sense must be increased to meet these needs.

## II. PSYCHOLOGICAL CONSERVATION

It would seem, then, that it is uneconomic, i.e., produces less satisfaction, to increase either the quantitative or qualitative levels of a person's desires, especially those which cannot be fulfilled. To illustrate, we may consider the case of a person who gets a great measure of satisfaction from his automobile which has an eighty horsepower motor and a top speed of eighty miles per hour. A new model is put out by the same company, a car which boasts one hundred horsepower and a maximum speed of ninety miles per hour. The car owner is exposed to these features of the new model with the result that he becomes dissatisfied with his own model although it continues to perform as well as formerly. This is a clear example of how raising a person's standard of quality causes him to derive less satisfaction from a good. If the car owner has sufficient purchasing power and buys the new model, the lost satisfaction will be restored.[3] However, should he not have sufficient purchasing power to buy the new model, his lost satisfaction will continue unless he can bring his

3. If he actually wanted to go faster than eighty miles per hour he might even exceed his former satisfaction. However, many "improvements" in goods may not be as important as the one in this illustration.

desires down to their former level. This will be difficult, perhaps impossible, for him to do.

In view of the foregoing, the expansion of wants in all its aspects cannot be accepted as economically productive and leading to greater satisfaction. Since the variety of human wants are infinitely expansible, economic satisfaction might be forever beyond the horizon no matter how rapidly physical production proceeded. Expanding the intensity of human wants is productive only if the goods are already acquired or about to be acquired. The latest fashions for women's clothing is a repeating example of the uneconomic creation of desires which will cause the present still serviceable wardrobes to give less satisfaction. Those women who cannot afford to replace their wardrobes cannot regain the economic satisfaction which they formerly had. The changing of desires so as to destroy the satisfying qualities of serviceable goods must be considered uneconomic activity. Therefore, when sellers cause the owners of goods to be dissatisfied with what they already have without presenting an economic good of any significant improvement, such action is obviously uneconomic.[4] This is particularly true if the dissatisfied owners do not have the purchasing power to acquire the new good. We are yet within the scope of economics for we are relating desires to scarce goods in order to see how these functions satisfy man. We are, however, recognizing human desires as a dynamic force which themselves can produce more or less satisfaction.

Psychological conservation may be defined as those activities which will preserve a person's economic satisfaction by maintaining the appreciation which he derives from the goods he has already acquired.[5] Thus, the usefulness of an acquired good can be prolonged by preventing psychological obsolescence. In the above illustrations of automobiles and clothing fashions, dissatisfaction with the goods that a person already has is coupled with and intensified by the creation of desires for slightly different products. Inasmuch as causing such a change in desires is uneconomic and at the same time exactly opposed to conservation, we might view those actions which maintain one's desires or appreciation for his wealth as actions of conservation which are also economic. However, such actions should not be classed as productive, because the usefulness of goods is only being *maintained* and not *increased*. I might emphasize here, never-

4. Such action by sellers is uneconomic for the additional and traditional reason that it diverts resources which may be used for want satisfying purposes.

5. Another way of defining psychological conservation is that it consists of those actions which preserve the subjective value which a person places upon his stock of goods.

theless, that psychological production and conservation are both economic in that they are so directed as to cause a person's total economic satisfaction to be as large as possible.   Social actions are psychologically productive if they influence desires so that economic satisfaction is increased, and social actions involve psychological conservation if they keep desires for acquired goods unchanged so that economic satisfaction derived from the goods will not decline.

<div align="right">ARTHUR PETER BECKER.</div>

UNIVERSITY OF WISCONSIN,
MILWAUKEE EXTENSION DIVISION

# RECENT PUBLICATIONS

ADAM, H. T. L'Organisation Européenne de Coopération Économique. Paris, France: Librairie Générale de Droit et de Jurisprudence. 1949. pp. 292.

ANGELOPOULOS, ANGELOS. L'État et la Prospérité Sociale. Paris, France: Librairie Générale de Droit et de Jurisprudence. 1949. pp. 198.

BARNES, JAMES A. Wealth of the American People. New York: Prentice-Hall, Inc. 1949. pp. 910. $5.75.

BECKMAN, THEODORE N., AND ENGLE, NATHANAEL H. Wholesaling: Principles and Practice. New York: The Ronald Press Company. 1949. pp. 746. $5.00.

BLODGETT, RALPH H. Comparative Economic Systems. New York: The Macmillan Company. 1949. pp. 892. $5.00. [Revised edition.]

BRODSKY, MICHEL, AND ROCHER, PIERRE. L'Économie Politique Mathématique. Paris, France: Librairie Générale de Droit et de Jurisprudence. 1949. pp. 365.

BROWN, LYNDON O. Marketing and Distribution Research. New York: The Ronald Press Company. 1949. pp. 612. $5.00.

CARANDE, RAMON. Carlos V. y sus Banqueros. Madrid, Spain: Sociedad de Estudios y Publicaciones. 1949. pp. 634.

DETURK, E. E. Freedom from Want: A Survey of the Possibilities of Meeting the World's Food Needs. Waltham, Massachusetts: The Chronica Botanica Company. 1949. pp. 283. $2.00.

EINAUDI, LUIGI. Lezioni di Politica Sociale. Italy: The Author. 1949. pp. 252.

FABRICANT, SOLOMON. The Rising Trend of Government Employment. New York: National Bureau of Economic Research. 1949. pp. 30.

GROSS, HERMAN. A Theory to Stabilize the Business Cycle. Providence, Rhode Island: The Author. 1949. pp. 19.

HANCOCK, W. K., AND GOWING, M. M. History of the Second World War, British War Economy. London: His Majesty's Stationers. 1949. pp. 583. $5.50. [Volume I, Official History of the War Series.]

HEINIG, KURT. Das Budget. Stockholm, Sweden: J. C. B. Mohr (Paul Siebeck). 1949. pp. 586.

HIGGINS, BENJAMIN H. Lombard Street in War and Reconstruction. New York: National Bureau of Economic Research. 1949. pp. 108. $1.00.

HOAGLAND, HENRY E. Real Estate Principles. New York: McGraw-Hill Book Company, Inc. 1949. pp. 667. $5.00. [2d edition.]

HUNT, PEARSON, AND WILLIAMS, CHARLES M. Case Problems in Finance. Chicago, Illinois: Richard D. Irwin, Inc. 1949. pp. 402. $4.50.

INTERNATIONAL MONETARY FUND. Balance of Payments Yearbook, 1938, 1946, 1947. Washington, D. C.: The Fund. 1949. pp. 383.

JACQUEMYNS, G. La Maison Heureuse. Brussels, Belgium: Institut Universitaire d'Information Sociale et Économique. 1949. pp. 48.

———. Les Budgets Familiaux d'Ouvriers et d'Employés, 1947–1948. Brussels, Belgium: Institut Universitaire d'Information Sociale et Économique. 1949. pp. 56.

———. Vacances et Congés Payés des Belges. Brussels, Belgium: Institut Universitaire d'Information Sociale et Économique. 1949. pp. 63.

JESNESS, O. B. Readings on Agricultural Policy. Philadelphia, Pa.: The Blakiston Company. 1949. pp. 470. $4.75.

LINTNER, JOHN. Corporate Profits in Perspective. American Enterprise Association, Inc. 1949. pp. 50. 50 cents. [No. 432 in the Series "National Economic Problems."]

McMILLAN, S. STERLING. Individual Firm Adjustments under O.P.A. Bloomington, Indiana: The Principia Press, Inc. 1949. pp. 256. $3.50.

MANNING, RAYMOND E. Federal Excise Taxes. Washington, D. C.: The Library of Congress. 1949. pp. 104. 70 cents. [Public Affairs Bulletin No. 72.]

MAYER, HANS. Hundert Jahre Osterreichischer Wirtschaftsentwicklung. Vienna, Austria: Springer. 1949. pp. 714.

MIKESELL, RAYMOND E., AND CHENERY, HOLLIS B. Arabian Oil: America's Stake in the Middle East. Chapel Hill, North Carolina: University of North Carolina Press. 1949. pp. 201. $3.50.

MILLS, FREDERICK C. Statistical Methods Applied to Economics and Business. New York: Henry Holt and Company. 1949. pp. 746.

MONTGOMERY, ROBERT H.; LENHART, NORMAN J.; AND JENNINGS, ALVIN R. Auditing. New York: The Ronald Press Company. 1949. pp. 690. $6.00. [7th edition.]

MOULTON, HAROLD GLENN. The State of the American Economy. Los Angeles: Haynes Foundation. 1949. pp. 70.

MUNSHI, M. C., AND KARNIK, K. P. Industrial Profits in India (1936–1944). (An Inductive Study.) New Delhi, India: Federation of Indian Chambers of Commerce & Industry. 1948. pp. 326.

PECK, GUSTAV. The Question of a Flexible Statutory Minimum Wage. Washington, D. C.: The Library of Congress. 1949. pp. 26. 25 cents. [Public Affairs Bulletin No. 73.]

RANDALL, FRANK A. History of the Development of Building Construction in Chicago. Urbana, Illinois: University of Illinois Press. 1949. pp. 388.

SANDERS, THOMAS H. Company Annual Reports to Stockholders, Employees, and the Public. Boston, Massachusetts: Harvard Business School. 1949. pp. 338. $3.75.

SOVANI, N. V. Post-War Inflation in India — A Survey. Poona, India: Gokhale Institute of Politics and Economics. 1949. pp. 94. Rs. 5.

TRAVERSIN, GUY. L'Agriculture en Grande-Bretagne, Tomes I and II. Paris, France: Institut National de la Statistique et des Études Économiques. 1949. pp. 476.

UGO PAPI, GIUSEPPE. Préliminaires à l'Étude des Plans pour l'Après-Guerre. Rome, Italy: Institut International d'Agriculture. 1946. pp. 306.

UNITED NATIONS, DEPARTMENT OF ECONOMIC AFFAIRS. Inflationary and Deflationary Tendencies, 1946–1948. Lake Success, New York: United Nations. 1949. pp. 52. 40 cents.

U. S. DEPARTMENT OF COMMERCE. Survey of University Business and Economic Research Projects. Washington, D. C.: The Department. 1947–48. pp. 240.

WEINTRAUB, SIDNEY. Price Theory. New York: Pitman Publishing Company. 1949. pp. 447. $5.50.

---

# American Economic Review

## CONTENTS

VOLUME XXXIX — SEPTEMBER, 1949